

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS BIOLÓGICAS

Departamento de Bioquímica y Biología Molecular I



TESIS DOCTORAL

**Aplicaciones del DNA antiguo a muestras humanas :
relaciones familiares y estudios poblacionales diacrónicos**

MEMORIA PARA OPTAR AL GRADO DE DOCTOR
PRESENTADA POR

Cristina Gamba

DIRECTORES:

Eduardo, dir Arroyo Pardo
Eva, dir Fernández Domínguez

Madrid, 2015



UNIVERSIDAD COMPLUTENSE



5327665921

T347.21
GAM
apl

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS BIOLÓGICAS

Departamento De Bioquímica y Biología Molecular I



2.º de Biología
Log. de B. M. I.
S.

**Aplicaciones del DNA antiguo a muestras
humanas: relaciones familiares y estudios
poblacionales diacrónicos**

627627895

TESIS DOCTORAL

Cristina Gamba



46964253

Madrid, 2011

Este trabajo de tesis se ha realizado gracias a la financiación de los siguientes proyectos de investigación y la siguiente beca predoctoral:

Proyectos:

- Estudio de la transición Mesolítico-Neolítico en la Cuenca Mediterránea mediante el análisis de muestras biológicas antiguas (CGL2009-07959). Ministerio de Ciencia e Innovación (2009-2011). Investigador Principal: Eduardo Arroyo Pardo.
- Estudio del poblamiento del noreste Peninsular mediante el análisis de la diversidad de ADN mitocondrial de muestras biológicas antiguas procedentes de yacimientos de época Calcolítica. Universidad Complutense de Madrid-Comunidad Autónoma de Madrid (2009-2010). Investigador Principal: Eva Fernández Domínguez.
- Estudio de la Neolitización de la Cuenca Mediterránea mediante análisis de la variabilidad genética de muestras biológicas antiguas (CGL2006-07828/BOS). Ministerio de Educación y Ciencia (2006-2009). Investigador Principal: Eduardo Arroyo Pardo.
- Identificación de las secuencias genéticas en los enterramientos múltiples de época Calcolítica en Castellón (ADN mitocondrial). Fundación Dávalos-Fletcher (Castellón) (2006-2007). Investigador principal: Arturo Oliver Foix.
- Análisis de polimorfismos genéticos autosómicos, de Cromosoma Y y de ADN mitocondrial en muestras biológicas (Art.83 L.O.U) (2005-actualidad). Investigador principal: Eduardo Arroyo Pardo.

El muestreo ha sido realizado gracias a los siguientes proyectos:

- The last hunter-gatherers and the first farming communities in the south of the Iberian Peninsula and north of Morocco: a socio-economic approach through the management of production instruments and exploitation of the domestic resources (I&D-PTDC/HAH/64548/2006). Fundação para a Ciência e a Tecnologia-FCT (Portugal) (2008-2011). Investigador Principal: Juan Francisco Gibaja Bao.
- El yacimiento prehistórico de Sant Pau del Camp. Instituto Cultura de Barcelona (Ayuntamiento de Barcelona) (2007-2008). Investigador principal: Miquel Molist Montaña.

Beca:

F.P.U. (Formación del Personal Universitario), AP2006-01586. Ministerio de Educación.

A mi familia

Navigia atque agri culturas moenia leges
arma vias vestes et cetera de genere horum,
praemia, delicias quoque vitae funditus omnis,
carmina, picturas et daedala signa polita
usus et impigrae simul experientia mentis
paulatim docuit pedetemptim progredientis.

Titus Lucretius Carus, De rerum natura (Liber V)

[Naves, cultivos de los campos, fortificaciones, leyes, /
armas, caminos, vestidos y demás beneficios de este tipo, /
también los goces de la vida completamente todos, /
poemas, pinturas y artísticas estatuas pulidas, / el uso y, al
mismo tiempo, la experiencia de su espíritu activo / poco
a poco enseñó a los hombres, que avanzaban paso a paso.

Tito Lucrecio Caro, La naturaleza de las cosas (Libro V)]

En este trabajo se han implicado muchas personas a las que querría agradecer.

En primer lugar, me gustaría agradecer a mis directores de tesis el Dr. Eduardo Arroyo Pardo y la Dra. Eva Fernández Domínguez por darme la posibilidad de realizar este trabajo de tesis en lo que me gusta. Muchas gracias por vuestras enseñanzas y vuestro apoyo constante durante estos años. Gracias por confiar en mí, animarme y creer en este trabajo.

En segundo lugar querría agradecer a mis tutores durante las estancias en el extranjero y a sus colaboradores. A Marie-France Deguilloux y Marie-Hélène Pemonge por su disponibilidad y ayuda durante mi estancia en su laboratorio. A Lounès Chikhi y Rita Rasteiro por introducirme en el complicado mundo de las simulaciones de manera sencilla. Gracias a todos ellos por querer mantener una colaboración estricta y confiar en este trabajo de tesis.

Este trabajo multidisciplinar no habría sido posible sin la confianza depositada por parte de los arqueólogos, antropólogos y especialistas que han cedido las muestras para su análisis genético y que han colaborado en la realización de este estudio y en la interpretación de los resultados. Muchas gracias por su colaboración a: Miquel Molist, Arturo Oliver, Pilar Utrilla, Vicente Baldellou, Manuel Edo, Anna Blasco, Pepa Villalba, Teresa Cabellos, Antonio Carvalho, Juan Gibaja, Eulalia Subirá, Jordi Ruiz, Váγγελis Villar, Francisco Pastor, Alejandro Pérez, Ferrán Estebaranz, Laura Martínez, Josep Anfruns, Mohaammad Alrousan, João Zilhão, Daniel Turbón, José Ramos, Ángel Esparza, Germán Delibes, José Antonio Sánchez, Didier Binder, Henry Duday, Carmen Olaria, María del Carmen Molestina, Chema López y Domingo Martínez.

Querría también agradecer a los equipos de investigación de Trinidad de Torres y de Terry Brown que han mantenido una colaboración con nuestro grupo de investigación a lo largo de estos años.

Gracias a mi tutor Antonio Tormo por su disponibilidad y por demostrar interés en este trabajo.

Gracias a Manuel Guzmán por su asesoramiento en la fase final de este trabajo de tesis.

Gracias a mis compañeros Carlos Baeza, Ana María López, Jorge Abarca, María Ruíz-Herrera, César Yoshi y Sara Palomo. Más especialmente a Mirian Tirado, por su contribución en la realización de este trabajo.

Gracias a todos los miembros del departamento de Antropología de la UCM, del Grupo de Genética de Poblaciones y Conservación del Instituto Gulbenkian (Oeiras, Portugal) y del laboratorio LAPP – PACEA de la Universidad de Burdeos 1 (Talence, Francia).

Gracias a todos los profesores que me han introducido en el mundo de la genética de poblaciones, especialmente a mis directoras del trabajo de fin de carrera: María Soledad Mesa, Patrizia Malaspina y Ana María López.

El ánimo y el apoyo de mis padres en esta aventura lejos de casa han sido fundamentales. Grazie davvero Ugo e Giusy.

Muchas gracias a mi hermana, por estar siempre pendiente de mí y ser también una gran amiga. Grazie Chiara!

Estos años en Madrid no habrían sido tan bonitos sin mis amigos siempre a mi lado. Muchas gracias a Bea, Manu, David, Pedro, María José, Pablo, Jesús, Cristina, Alberto, Diego, María y David por apoyarme y compartir supermartes, fiestas, cumpleaños, conciertos, o unas simples cañas conmigo. Un agradecimiento especial a mi directora Eva, por su sincera amistad.

Fuera de Madrid, gracias a mis amigos de Bergamo, especialmente a Chiara y Guido, a mi querida amiga de Roma, Anna, a Carolina por hacerme sentir en mi casa en Oeiras y a Giuseppe, en Burdeos.

Y dejo por último el que en mi vida es el primero. Muchas gracias Juan, por estar siempre a mi lado.

ÍNDICE GENERAL

ÍNDICE GENERAL	i
ÍNDICE DE FIGURAS	vi
ÍNDICE DE TABLAS	ix
ABREVIATURAS	xi
RESUMEN (Español)	1
ABSTRACT (English)	3
INTRODUCCIÓN	7
I.1 Justificación del trabajo	7
I.2 Marcadores genéticos en el estudio del ADN antiguo	9
I.2.1 El ADN mitocondrial	10
I.2.1.1 Características del ADN mitocondrial	12
I.2.1.2 Variabilidad genética del ADN mitocondrial	14
I.2.2 El ADN nuclear	17
I.2.2.1 Marcadores autosómicos	18
I.2.2.1.1 El cromosoma Y	19
I.2.2.1.2 Variabilidad genética del Cromosoma Y	19
I.2.2.2 El cromosoma X	22
I.3 El ADN antiguo: historia y problemáticas asociadas a su análisis	24
I.3.1 Historia del ADN antiguo	24
I.3.2 Factores de degradación del ADN antiguo	27
I.3.2.1 Degradación enzimática e hidrólisis	28
I.3.2.2 Oxidación	32
I.3.2.3 Puentes cruzados o "crosslinks"	33
I.3.3 Problemáticas de los estudios de ADN antiguo	33
I.3.3.1 Jumping PCR	33
I.3.3.2 Inhibidores de la reacción de PCR	34
I.3.3.3 Contaminación y criterios de autenticidad	34

I.3.3.4 Influencia de los factores ambientales y del entorno de deposición sobre la preservación del ADN antiguo	41
I.3.3.5 Límite de preservación del ADN antiguo	42
I.4 Aplicaciones del ADN antiguo	44
I.4.1 Estudios diacrónicos en Europa y la Península Ibérica, desde el Neolítico hasta la Edad del Hierro	44
I.4.1.1 Marco histórico	44
I.4.1.1.1 El Neolítico	47
I.4.1.1.2 Transición a la Edad de los Metales y la Edad del Cobre	52
I.4.1.1.3 Edad del Bronce	55
I.4.1.1.4 Edad del Hierro	57
I.4.1.2 Datos genéticos de la población moderna	59
I.4.1.3 Datos genéticos de las poblaciones antiguas	69
I.4.2 Relaciones familiares e Identificación	74
OBJETIVOS (Español)	83
AIMS AND SCOPE (English)	87
ARTÍCULOS / ARTICLES	91
ARTÍCULO 1 / ARTICLE 1	93
Resumen (Español)	93
Anexo - ARTÍCULO 1	97
Appendix - ARTICLE 1	98
ARTÍCULO 2 / ARTICLE 2	99
Resumen (Español)	99
ARTÍCULO 3 / ARTICLE 3	103
Resumen (Español)	103
Supplementary Material	117
Table S1	117
Table S2	118
Table S3	119
Table S4	122
Table S5	123
Table S6	124
Table S7	125
Figure S1	126

ARTÍCULO 4 / ARTICLE 4	127
Resumen (Español)	127
Actualización - ARTÍCULO 4	133
Tablas	134
Anexo I	135
Anexo II	138
Update - ARTICLE 4	139
Tables	140
Appendix I	141
Appendix II	144
ARTÍCULO 5 / ARTICLE 5	147
Resumen (Español)	147
Supplementary Material	157
Figure S1	158
Samples pictures	159
Table S1	160
Figure S2	161
Table S2	161
DISCUSIÓN	165
Consideraciones generales	165
DISCUSIÓN DE LOS OBJETIVOS – METODOLOGÍA Y VALIDACIÓN DE LOS RESULTADOS	166
D.1 Comparación de diferentes metodologías disponibles en la literatura de ADN antiguo y forense para el aislamiento y la amplificación de ADN de muestras arqueológicas. Selección de aquellos protocolos más eficientes y que se ajustan a las características de las muestras disponibles y a las finalidades concretas del estudio.	166
D.1.1 Comparación de diferentes metodologías para la extracción de ADN (ARTÍCULO 1)	167
D.1.2 Comparación de diferentes metodologías para la amplificación de ADN (ARTÍCULOS 2-5)	173
D.1.2.1 Amplificación del ADN nuclear mediante el <i>kit MiniFiler</i>	176
D.1.2.2 Amplificación del ADN mitocondrial mediante el <i>kit Multiplex</i> de <i>Qiagen</i>	178
D.1.3 Selección de las muestras y de la metodología adecuada en función de las finalidades del estudio (ARTÍCULOS 3-5)	181
D.2 Evaluación de la posibilidad de recuperar ADN a partir de muestras antiguas en función de diferentes variables como son el estado de preservación, la procedencia, la antigüedad, las	

características morfológicas del tejido y el contenido de ADN en los extractos estimado mediante métodos espectrofotométricos.	187
D.2.1 Procedencia y antigüedad de las muestras (Artículos 1, 3 y 4)	187
D.2.2 Características de las muestras (Artículos 1 y 5)	189
D.2.3 Evaluación de las medidas de absorbancia (Artículo 1)	192
D.3. Validación de los resultados de ADN mitocondrial y nuclear obtenidos a partir de muestras históricas y prehistóricas mediante los pertinentes criterios de autenticidad, aceptados internacionalmente.	195
D.3.1 Detección y control de la contaminación durante la extracción y la amplificación (Artículos 3-5)	195
D.3.1.1 Contaminación en los blancos de extracción	196
D.3.1.2 Contaminación en los blancos de PCR	197
D.3.1.3 Contaminación no detectada en los blancos	199
D.3.2 Replicación de los análisis en un segundo laboratorio	204
D.3.3 Consideraciones generales sobre la validación de los resultados	207
DISCUSIÓN DE LOS OBJETIVOS ESPECÍFICOS – ESTUDIO DIACRÓNICO-POBLACIONAL	209
D.4 Estimación de la continuidad/discontinuidad genética entre los pobladores prehistóricos (neolíticos, calcolíticos, de la Edad del Bronce e íberos) y actuales del noreste de la Península Ibérica (Comunidad Valenciana, Cataluña y Aragón) y evaluación del impacto genético del Neolítico en esta misma área.	209
D.4.1 Comparación de la composición de haplotipos y haplogrupos de las poblaciones neolíticas, calcolíticas, de la edad del Bronce e íberas del noreste peninsular entre sí y con otras poblaciones, antiguas y actuales, de la misma región (Artículos 3 y 4)	209
D.4.1.1 Composición haplotípica de las poblaciones	213
D.4.1.2 Composición de haplogrupos de las poblaciones	215
D.4.2 Utilización de métodos filogenéticos, estadísticos y computacionales para la interpretación de los resultados, que incluyen la estimación de la continuidad/discontinuidad genética entre diferentes períodos y la evaluación del impacto del Neolítico en la región estudiada (Artículos 3 y 4)	225
D.4.2.1 Comparación entre distancias genéticas observadas y simuladas	228
D.4.2.2 Selección del modelo más adecuado	236
DISCUSIÓN DE LOS OBJETIVOS – RELACIONES FAMILIARES	239
D.5 Verificación de las hipótesis de parentesco y de determinación del sexo formuladas a partir de otras disciplinas (arqueología, historiografía, genealogía, antropología) mediante el análisis	

genético de marcadores nucleares STR en un conjunto de muestras medievales procedentes de sepulturas múltiples (Iglesia de San Esteban de Cuéllar, Segovia).	239
D.5.1 Interpretación de los perfiles STRs (Artículo 5)	239
D.5.2 Interpretación de los resultados genéticos obtenidos a la luz de hipótesis arqueológicas, valorándolos conjuntamente con la información genealógica, histórica, arqueológica y antropológica disponible (Artículo 5)	243
D.5.2.1 Determinación molecular del sexo	244
D.5.2.2 Relaciones familiares	245
CONCLUSIONES (Español)	251
CONCLUSIONS (English)	259
BIBLIOGRAFÍA	267
CD	291

ÍNDICE DE FIGURAS

Figura I 1 Esquema de la molécula de mtDNA. Se detallan la localización del D-Loop y, en su interior, las regiones hipervariables. También se representa la localización de los principales genes a lo largo de la molécula y las mutaciones asociadas a los principales haplogrupos. _____	12
Figura I 2 Relación filogenética entre los principales haplogrupos mitocondriales en correspondencia de las regiones geográficas (imagen modificada a partir de http://mitomap.org). _____	16
Figura I 3 Árbol filogenético de los haplogrupos del cromosoma Y (imagen de Jobling y Tyler-Smith, 2003). _____	21
Figura I 4 Principales modificaciones post-mortem del ADN. Las flechas rojas indican los principales sitios de depuración, las azules los de daño oxidativo y las verdes los de hidrólisis (imagen modificada a partir de Hofreiter et al. 2001). _____	29
Figura I 5 Mecanismos de actuación de la desaminación hidrolítica. Se detallan las lesiones de tipo 1 y de tipo 2. _____	31
Figura I 6 Mutaciones asociadas a la desaminación hidrolítica de tipo 1 y 2. Se ilustran los distintos resultados dependiendo de si la Miscoding Lesion tiene lugar en la cadena ligera (C_L) o en la pesada (C_H). _____	31
Figura I 7 Esquema de los posibles movimientos poblacionales que tuvieron lugar en Europa durante la expansión del hombre moderno (en rojo), la repoblación de Europa a partir de los refugios paleolíticos (en morado) y la difusión de la agricultura (azul). Las flechas se refieren a los posibles movimientos poblacionales/culturales sugeridos por los datos arqueológicos, y los óvalos se corresponden con el comienzo de cada movimiento. Se señalan con flechas bidireccionales los movimientos de contracción demográfica en los refugios paleolíticos (óvalos morados), seguidos de una expansión poblacional después del último máximo glacial. La flecha discontinua representa la expansión post-glacial desde el refugio de la franja cantábrica hacia Europa septentrional (Barbujani y Goldstein, 2004). _____	46
Figura I 8 Mapa en la que se representan las dataciones de las diferentes culturas neolíticas, expresadas en años a.C. (imagen modificada a partir de Guilaine 2003). _____	49
Figura I 9 a) Distribución de los monumentos megalíticos en Europa; b) Localización de algunas de las culturas de la Edad del Cobre y del Bronce: cultura del vaso campaniforme, de ánforas globulares, de la cerámica cordada y del hacha de combate; c) Difusión de la cultura de las culturas herederas de los Campos de Urnas (Urnfield culture) durante el primer milenio AP; d) Mapa de Europa durante el primer milenio a.C.: las influencias célticas y orientales (griegas y fenicias) (imágenes modificadas a partir de Cavalli-Sforza et al. 1994) _____	54

Figura I 10 Mapas sintéticos de la Península Ibérica en la que se representan las primeras tres componentes principales del estudio de Bertranpetit y Cavalli-Sforza (1991): a) primera componente principal, asociada a un 27 % de la variabilidad total, b) segunda componente principal, asociada a un 14 % de la variabilidad total, c) tercera componente principal, asociada a un 12 % de la variabilidad total. El intervalo entre el valor máximo y el mínimo de cada componente principal se ha subdividido en seis rangos iguales. La dirección de crecimiento de los valores de cada componente principal fue elegida de manera arbitraria (imagen modificada a partir de Bertranpetit y Cavalli-Sforza 1991).	60
Figura I 11 Datación molecular de los principales haplogrupos europeos (imagen modificada a partir de Richards et al. 2000).	65

Figura D 1 (a) Eficiencia en la amplificación ("Eficiencia PCR") frente a los diferentes yacimientos arqueológicos analizados, ordenados de la letra "a" a la "g" en orden cronológico de más antiguo a más reciente. (b) Eficiencia en la obtención de resultados replicables (media de la "Eficiencia 3,4" y "Eficiencia 5,6") frente al método de extracción. (c) Poder inhibidor de los extractos en función del método de extracción empleado (N.I.=Sin Inhibición, P.I.=Inhibición Parcial, T.I.=Inhibición Total). (d) Poder de inhibición frente a los dos métodos de extracción basados en el empleo de sílica, uno incluye GuSCN y otro NaCl. P/C: protocolo basado en el empleo de fenol/cloroformo (Fernández 2005); S-B: protocolo basado en el empleo de sílica (Rohland y Hofreiter 2009); Q: protocolo QIAamp DNA Micro Kit (Qiagen) (Figura 1 del ARTÍCULO 1).	168
--	-----

Figura D 2 Gráfico de la distribución de haplogrupos en las poblaciones antiguas con las que se han comparado los datos de la presente tesis. Referencias bibliográficas: Calcolítico Centro Europa (CWC) [Haak et al. 2008], Íberos Girona, España [Sampietro et al. 2005], Neolítico Antiguo Centro Europa (LBK) [Haak et al. 2005, 2010], Mesolítico Centro Europa [Bramanti et al. 2009], Neolítico Final Francia [Lacan et al. 2011], Neolítico Medio Barcelona, España [Sampietro et al. 2007], Neolítico Medio Francia [Deguilloux et al. 2011], Mesolítico Norte de Europa [Malmström et al. 2009], Neolítico Norte de Europa [Malmström et al. 2009].	217
---	-----

Figura D 3 Probabilidad de obtener F_{ST} simulados mayores que los F_{ST} observados entre parejas de poblaciones en los tres modelos explorados (TMP, SM, SDGM). N_e : tamaño efectivo poblacional. N_e -Neo: rango de N_e . El color blanco se refiere a valores superiores a 0.05 (5 %). Gráficos incluidos en los ARTÍCULOS 3 y 4.	231
--	-----

Figura D 4 Probabilidad de obtener F_{ST} simulados mayores que los F_{ST} observados entre parejas de poblaciones, en correspondencia de los tres modelos explorados (TMP, SM, SDGM). N_e : tamaño efectivo poblacional. N_e -Neo: rango de N_e . El color blanco se refiere a valores superiores a 0.05 (5 %). Gráficos incluidos en el ARTÍCULO 3 y 4.	235
--	-----

ÍNDICE DE TABLAS

Tabla D 1 Porcentaje de éxito en el análisis de las muestras validadas en los ARTÍCULOS 3, 4 y 5 en función del período analizado. Se aporta también el porcentaje de inhibición de los extractos en los yacimientos en los que fue testado. Debajo de cada período se proporciona la media del éxito de amplificación de los diferentes yacimientos incluidos y los marcadores analizados. _____	175
Tabla D 2 Porcentajes de amplificaciones contaminadas en relación a los períodos estudiados, el yacimiento, y el tipo de marcadores analizados. Se señala también el porcentaje de éxito en la obtención y validación de los resultados. _____	199
Tabla D 3 Porcentajes de muestras que no proporcionaron resultados reproducibles por yacimiento y período cronológico analizado. Se distingue entre las diferentes causas de fracaso: la ausencia de amplificación y la detección de contaminantes en el análisis de secuencias directas y clones, señalándose la fuente de contaminación más probable. _____	199
Tabla D 4 Relación del número de muestras analizadas en un segundo laboratorio. Se detallan los porcentajes de replicación exitosos así como las causas de la ausencia de replicación. _____	206
Tabla D 5 Tabla resumen de los resultados obtenidos en los ARTÍCULOS 3 y 4 (respectivamente Tabla A1 y Tabla 1). Los haplotipos coincidentes están marcados con el mismo color (dentro de esta tabla y con la Tabla D 6). _____	212
Tabla D 6 Haplotipos compartidos entre las muestras estudiadas y las disponibles en la literatura. Se marcan las coincidencias de haplotipos con los mismos colores empleados en la Tabla D 5 . _____	213
Tabla D 7 Distancias genéticas (F_{ST}) entre las poblaciones analizadas (Tabla A2 del Anexo I del ARTÍCULO 4). _____	227
Tabla D 8 Resumen de los resultados obtenidos en las comparaciones entre F_{ST} observadas y simuladas, elaborado a partir de los gráficos presentados en el presente apartado. Se especifican por cada pareja de poblaciones comparadas para qué valores es posible detectar F_{ST} simuladas mayores que las observadas en al menos el 5 % de los casos. _____	229
Tabla D 9 Probabilidad de identificar el modelo de procedencia seleccionando 500 valores al azar dentro del conjunto de datos simulados (1 millón) por cada uno de los modelos: TPM (primera línea), SM (segunda línea) y SDGM (tercera línea) (Tabla S4 del ARTÍCULO 3). _____	237
Tabla D 10 Perfiles de las muestras amplificadas con el kit MiniFiler. Los alelos con un asterisco (en las celdas grises) se refieren al fenómeno del allelic dropout y la presencia de un guión, a la ausencia de resultados (locus drop-out). _____	242

ABREVIATURAS

A	<u>A</u> denina
ABC	método computacional de aproximación bayesiana, en inglés <u>A</u> pproximate <u>B</u> ayesian <u>C</u> omputation
a.C.	antes de <u>C</u> risto
aDNA	ADN antiguo, en inglés <u>a</u> ncient <u>D</u> N <u>A</u>
AP	<u>A</u> ntes del <u>P</u> resente
C	<u>C</u> itosina
CNVs	variantes de número de copias, en inglés <u>C</u> opy <u>N</u> umber <u>V</u> ariants
CRS	secuencia de referencia de Cambridge, en inglés <u>C</u> ambridge <u>R</u> eference <u>S</u> equ <u>e</u> nce,
D-Loop	Lazo D o región control del mtDNA, en inglés <u>D</u> isplacement <u>L</u> oop o <u>C</u> ontrol <u>R</u> egion
EUP	Paleolítico Superior Inicial, en inglés <u>E</u> arly <u>U</u> pper <u>P</u> alaeolithic
G	<u>G</u> uanina
GuSCN	<u>T</u> iocianato de <u>G</u> uanidina
H	cadena pesada del ADN mitocondrial, en inglés <u>H</u> eavy
HAM	<u>H</u> ombre <u>A</u> natómicamente <u>M</u> oderno
HLA	antígenos leucocitarios humanos, en inglés <u>H</u> uman <u>L</u> eukocyte <u>A</u> ntigens
HVR	Regiones Hipervariables del mtDNA, en inglés <u>H</u> ypervariable <u>R</u> egions
HVS	Secuencias Hipervariables del mtDNA, en inglés <u>H</u> ypervariable <u>S</u> equences
kb	<u>k</u> ilo <u>b</u> ases, miles de bases
km	<u>K</u> iló <u>m</u> etro
L	cadena ligera del ADN mitocondrial, en inglés <u>L</u> ight
LD	desequilibrio de ligamiento, en inglés <u>L</u> inkage <u>D</u> isequilibrium
LTDNA	bajo número de moléculas de ADN molde, en inglés <u>L</u> ow <u>T</u> emplate <u>D</u> N <u>A</u>
LGM	último máximo glacial, en inglés <u>L</u> ast <u>G</u> lacial <u>M</u> aximum
LUP	Paleolítico Superior Final, en inglés <u>L</u> ate <u>U</u> pper <u>P</u> alaeolithic
Mb	<u>M</u> ega <u>b</u> ases, millones de bases
mg	<u>m</u> iligramos
ml	<u>m</u> ililitros
mtDNA	ADN mitocondrial, en inglés <u>m</u> itochondrial <u>D</u> N <u>A</u>
MUP	Paleolítico Superior Medio, en inglés <u>M</u> iddle <u>U</u> pper <u>P</u> alaeolithic
μl	<u>m</u> icrolitro

NaCl	<i>Cloruro Sódico</i>
N_e	tamaño poblacional efectivo
NGS	técnicas de secuenciación de nueva generación, en inglés <i><u>N</u>ext <u>G</u>eneration <u>S</u>equencing</i>
nm	<i><u>N</u>anó<u>m</u>etros</i>
NRY	región no recombinante del cromosoma Y, en inglés <i><u>N</u>on <u>R</u>ecombinig region of the <u>Y</u> chromosome</i>
nuDNA	ADN nuclear, en inglés <i><u>n</u>uclear <u>D</u>NA</i>
NUMTs	pseudogenes del ADN mitocondrial insertados en el ADN Nuclear, en inglés <i><u>N</u>uclear <u>m</u>itochondrial <u>D</u>NA</i>
PAR	región pseudoautosómica, en inglés <i><u>P</u>seudo-<u>A</u>utosomic <u>R</u>egion</i>
pb	pares de bases, en inglés <i>base pairs</i>
P/C	protocolo de extracción de ADN que emplea Fenol Cloroformo, en inglés <i><u>P</u>henol/<u>C</u>hloroform</i>
PCA	análisis de las componentes principales, en inglés <i><u>P</u>rincipal <u>C</u>omponent <u>A</u>nalysis</i>
PCR	reacción en cadena de la polimerasa, en inglés <i><u>P</u>olymerase <u>C</u>hain <u>R</u>eaction</i>
Q	protocolo de extracción de ADN basado en el empleo de kits comerciales de la casa <i><u>Q</u>iagen</i> .
rCRS	secuencia de referencia de Cambridge revisada, en inglés <i><u>r</u>evised <u>C</u>ambridge <u>R</u>eference <u>S</u>equences</i>
RFLPs	polimorfismos de longitud de los fragmentos de restricción, en inglés <i><u>R</u>estriction <u>F</u>ragment <u>L</u>ength <u>P</u>olymorphisms</i>
S-B	protocolo de extracción de ADN basado en el empleo de sílica, en inglés <i><u>S</u>ilica-<u>B</u>ased</i>
SNP	polimorfismo de un solo nucleótido, en inglés <i><u>S</u>ingle <u>N</u>ucleotyde <u>P</u>olymorphism</i>
SRY	región del cromosoma Y que determina el sexo masculino, en inglés <i><u>S</u>ex-determinig <u>R</u>egion of <u>Y</u></i>
STRs	secuencias cortas repetidas en tándem, en inglés <i><u>S</u>hort <u>T</u>andem <u>R</u>epeats</i>
T	<i>T</i> imina
TMRCa	tiempo de coalescencia respecto al ancestro común más reciente, en inglés <i><u>T</u>ime to the <u>M</u>ost <u>R</u>ecent <u>C</u>ommon <u>A</u>ncestor</i>
UEP	evento polimórfico único, en inglés <i><u>U</u>nique <u>E</u>vent <u>P</u>olymorphisms</i>
YCC	consorcio del cromosoma Y, en inglés <i><u>Y</u> <u>C</u>hromosome <u>C</u>onsortium</i>

RESUMEN (Español)

El ADN antiguo es una disciplina de reciente incorporación al campo científico y su desarrollo ha sido posible gracias a los avances de las técnicas de genética molecular y forense. El análisis de marcadores genéticos de muestras humanas antiguas ofrece una oportunidad única para investigar y comprobar hipótesis planteadas desde la arqueología, la historiografía, la antropología o la genética de poblaciones. Sin embargo, está dificultada por la escasez de muestras, la degradación del material genético y los problemas de contaminación.

En la presente tesis doctoral se abordan aspectos metodológicos y aplicaciones de esta disciplina a restos humanos antiguos de épocas comprendidas entre el Neolítico y la Edad Media.

En lo que se refiere a la implementación metodológica, la comparación de diferentes métodos de extracción descritos en la literatura ha permitido identificar como protocolo más eficiente el de Rohland y Hofreiter (2007), basado en el empleo de sílica. Este protocolo ha venido a sustituir a aquel habitualmente empleado por nuestro grupo basado en el empleo de fenol y cloroformo. Las principales ventajas de dicho método consisten en su elevada eficiencia en la obtención de ADN y en la reducción de la manipulación y consecuentemente del riesgo de contaminación de las muestras con ADN actual. Por otra parte, en este protocolo no se produce una exposición a sustancias tóxicas volátiles, como el fenol y el cloroformo.

La evaluación de la influencia de diferentes variables en la obtención de resultados ha puesto de manifiesto que la procedencia y el entorno de disposición de las muestras son determinantes a nivel de su preservación molecular. Sin embargo, su antigüedad y características morfológicas no parecen jugar un papel tan importante.

Gracias a la citada experimentación metodológica y a la validación de los resultados, según criterios establecidos internacionalmente, ha sido posible llevar a cabo un estudio poblacional diacrónico y establecer relaciones familiares entre varios individuos procedentes de enterramientos múltiples.

Así, el análisis del ADN mitocondrial de muestras neolíticas, calcolíticas, de la Edad del Bronce e íberas procedentes del noreste de la península Ibérica ha permitido evaluar el poblamiento de esta región desde un punto de vista diacrónico. Mediante la implementación de los resultados obtenidos en simulaciones bayesianas coalescentes ha sido posible destacar los modelos demográficos que más se ajustan a los resultados genéticos obtenidos. Se ha demostrado que los escenarios panmícticos no permiten explicar los resultados genéticos obtenidos, mientras que los modelos estructurados muestran un mejor ajuste. Entre las conclusiones más relevantes merece la pena destacar que los resultados apuntan a una colonización pionera de pequeños grupos neolíticos en el noreste de la Península Ibérica. Posteriormente, la deriva genética pudo jugar un papel importante, puesto que los períodos sucesivos (Neolítico Medio, Calcolítico, Edad del Bronce y población actual) apuntan a la presencia de cierta continuidad genética hasta la actualidad, exceptuando el conjunto de muestras íberas, que presenta cierta diferenciación respecto a todas las demás poblaciones.

Por otra parte, se ha empleado una herramienta forense (*AmpF ℓ STR[®] MiniFiler[™] PCR Amplification Kit, Applied Biosystems*) para el análisis de ADN nuclear de siete individuos medievales enterrados en un conjunto de sepulcros (Iglesia de San Esteban de Cuellar, Segovia). De esta manera ha sido posible asignar el sexo a todos los individuos, rectificando la determinación realizada mediante estudios antropométricos en uno de ellos. Merece la pena destacar también que el análisis de los datos con un programa informático de análisis de *pedrigrees*, usualmente empleado en el campo forense, ha permitido excluir la hipótesis de paternidad entre dos individuos infantiles enterrados junto con un individuo adulto. Por otra parte, ha sido posible establecer una relación paterno-filial asociada a una elevada probabilidad (99,98 %) entre otro espécimen y el individuo adulto mencionado, confirmando una hipótesis arqueológica alternativa basada en inscripciones sepulcrales poco claras.

ABSTRACT (English)

Ancient DNA is a young field that has only recently joined the scientific community and has undergone rapid advancement thanks to technical innovations in molecular and forensic genetics. The analysis of genetic markers from ancient human samples provides us with a unique opportunity to test hypotheses from the fields of Archaeology, History, Anthropology and Population Genetics. However, sample availability, degradation of genetic material and contamination problems still represent a major issue.

The present PhD thesis explores both the methodological approaches and the applications of ancient DNA studies on human samples spanning from the Neolithic to Middle Ages.

A comparison of different DNA extraction methods described in the literature revealed that Rohland and Hofreiter's (2007) silica-based protocol was the most efficient methodology. This protocol replaced the Phenol/Chloroform DNA extraction method, commonly used in our laboratory. The major advantages of Rohland and Hofreiter's method are that it is highly efficient in recovering DNA and it reduces the handling of the sample, which leads to lower levels of contamination. Moreover, in this protocol there is no exposure to toxic volatile reagents such as Phenol and Chloroform.

A statistical evaluation of the different variables which can influence the results shows that the geographical origin and depositional environment of the sample play an important role in its molecular preservation. In contrast, sample age and morphology do not show a significant influence on DNA integrity.

Thanks to this research on methodologies and further validation of results based on internationally established criteria, it has been possible to carry out diachronic population studies and to establish kinship among individuals coming from multiple burials.

The settlement of North Eastern Iberia has been explored with a diachronic analysis of mitochondrial DNA from Neolithic, Chalcolithic, Bronze Age and Iberian samples. The implementation of Bayesian coalescent simulations has allowed us to determine which

demographic models show a better fit with the observed data. It has been demonstrated that panmictic scenarios could not explain the obtained genetic results, while structured models showed a better fit.

The conclusions contain notable findings such as an indication that the first colonisation was by small Neolithic groups in North Eastern Iberia. Afterwards, genetic drift probably played an important role, since comparisons among subsequent periods (Middle Neolithic, Chalcolithic, Bronze Age and the modern population) showed a certain degree of genetic continuity until the present. However, Iberian samples exhibited significant differences when compared to all other populations.

A forensic tool (*AmpFℓSTR® MiniFiler™ PCR Amplification Kit, Applied Biosystems*) has also been used to analyse nuclear DNA from seven medieval individuals buried in four tombs (San Esteban church, Cuellar, Segovia, Spain). With this technique it has been possible to establish the sex of all the individuals, one of which had been previously determined incorrectly using anthropometric methods. It is also worth noting that data analysis using forensic software for the analysis of pedigrees, made it possible to reject the hypothesis of paternity between an adult and two infants buried together. Moreover, this adult showed a significant probability (99.98 %) of being paternally related to another individual buried in a different tomb, as some burial inscriptions suggested.

INTRODUCCIÓN

INTRODUCCIÓN

I.1 JUSTIFICACIÓN DEL TRABAJO

Desde sus comienzos, el ADN antiguo (en inglés *ancient DNA*, aDNA) despertó el interés tanto de la comunidad científica como del público general. A principios de los años noventa diferentes grupos de investigación publicaron resultados muy impactantes a partir de muestras biológicas de millones de años de antigüedad (Golenberg et al., 1990; Woodward et al., 1994). Estos hallazgos inspiraron películas y libros en los que el alcance de esta disciplina parecía no tener límite. Sin embargo, en el curso de unos pocos años, se demostró que el ADN se degrada desde el momento de la muerte y que en condiciones óptimas se puede conservar, si bien muy dañado, durante un máximo de 100.000 años (Lindahl, 1993, 1997), permitiendo una recuperación de información genética a menudo escasa pero de valor único.

En el caso de la especie humana el ADN mitocondrial ha sido objeto de estudio en trabajos centrados en la evolución de los homínidos, las poblaciones humanas, su historia y sus relaciones filogenéticas. Este marcador se ha convertido en una herramienta de referencia en el estudio de linajes, debido a que se transmite a lo largo de la línea materna sin recombinación y tiene una tasa de mutación lo suficientemente elevada, como para adecuarse al estudio de la evolución reciente del hombre. Existen diferentes trabajos que emplearon el ADN mitocondrial para detectar el origen y dispersión del hombre moderno. Entre ellos merece la pena destacar el que realizaron Cann y colaboradores (Cann et al., 1987), que llegaron a identificar una “Eva mitocondrial”, es decir, una variante de ADN mitocondrial que se originó hace unos 200.000 años en África, aportando nuevos datos que apoyaban la hipótesis de un único origen africano del hombre moderno. Ésta hipótesis fue conocida como “out of

Africa” y, aunque se trata de la teoría actualmente más aceptada, ha sido objeto de un largo debate.

Otro caso muy debatido es el origen de la población europea. Se ha pretendido encontrar un consenso en la cuantificación del impacto genético de los pobladores neolíticos y paleolíticos. En este caso, en función del método de análisis de datos y del marcador genético empleado (ADN mitocondrial, cromosoma Y o marcadores autosómicos), se obtuvieron resultados discordantes (Richards, 2003; Dupanloup et al., 2004; Barbujani y Chikhi, 2006; Soares et al., 2010).

Gracias a las mejoras de las técnicas disponibles y su implementación por parte de equipos especializados, el ADN antiguo se ha convertido en una herramienta muy útil para proporcionar información acerca de cuestiones de diferente índole, que incluyen las arriba mencionadas, así como aplicaciones más específicas y concretas, como la determinación del sexo y estudios de identificación y relaciones familiares.

La presente tesis doctoral pretende aportar información acerca de la implementación y mejora de técnicas en el campo del ADN antiguo así como profundizar en diferentes cuestiones poblacionales, tales como el análisis diacrónico de la población del Levante Español desde el Neolítico hasta la época íbera y la determinación molecular del sexo y la estimación de relaciones familiares en un conjunto de sepulcros medievales.

I.2 MARCADORES GENÉTICOS EN EL ESTUDIO DEL ADN ANTIGUO

Los marcadores genéticos representan una herramienta muy interesante puesto que permiten estudiar la evolución de los seres vivos y de las poblaciones. Esto resulta posible gracias a la acumulación de cambios aleatorios o mutaciones en el ADN a lo largo del tiempo. El análisis de estas mutaciones en diferentes individuos, poblaciones o especies permite establecer relaciones de parentesco o filogenéticas en función de las muestras empleadas y de los marcadores estudiados. Las posibles aplicaciones abarcan diferentes áreas de la biología, como la genética forense, los estudios de ADN antiguo, los análisis genéticos de asociación con enfermedades, la genética de poblaciones y los estudios evolutivos y de conservación de especies. La evolución de las mutaciones que se detectan a nivel de estos marcadores se debe principalmente al azar y se van acumulando con el tiempo. En función de la aplicación concreta, se pueden emplear marcadores de linajes maternos o paternos, que en el caso de la especie humana se sitúan respectivamente en el ADN mitocondrial y en el cromosoma Y, o bien secuencias situadas en otros cromosomas.

Durante los últimos veinte años se ha ampliado muchísimo la información genética disponible acerca del origen del hombre y de las características y relaciones entre las diferentes poblaciones humanas. En el caso de muestras antiguas, el marcador genético más empleado es el ADN mitocondrial en especies animales y también el ADN presente en los cloroplastos en especies vegetales (Herrmann, 1994). La razón principal de esta elección se debe a la abundancia de estas moléculas en la célula, que supera la del ADN nuclear en unos cuatro o cinco órdenes de magnitud. Debido a la fragmentación y degradación del material genético con el tiempo, su abundancia asegura mayores probabilidades de recuperación y análisis.

1.2.1 El ADN mitocondrial

El ADN mitocondrial (en inglés *mitochondrial DNA*, mtDNA) humano es una molécula circular de doble cadena con una longitud de 16569 pb (pares de bases, en inglés *base pairs*) que se encuentra en el interior de las mitocondrias, orgánulos celulares encargados de la respiración celular. Se estima que hace unos 2 billones de años las mitocondrias establecieron una relación simbiótica con las células procariotas, cediendo posteriormente la mayoría de sus genes al núcleo (Wallin, 1923). En la actualidad el mtDNA contiene la información genética para la síntesis de 13 polipéptidos, todos ellos implicados en la fosforilación oxidativa, 2 RNAs ribosomales y 22 RNAs de transferencia. Estos últimos proporcionan los componentes necesarios para la traducción intramitocondrial. Sin embargo, la contribución genética mitocondrial a la fosforilación oxidativa es muy reducida comparada con la del ADN nuclear, cuyas proteínas esenciales son transportadas en el interior de las mitocondrias a través de su membrana (Shoubridge, 2001).

El mtDNA contiene varias regiones no codificantes, de las cuales la que presenta mayor tamaño se denomina Región de Control (*Control Region* o D-Loop, *Displacement Loop*) y tiene una longitud de 1122 pb (ver **Figura I 1**). Esta región está implicada en la regulación de la transcripción y de la replicación. En particular aquí se sitúan el origen de replicación de la cadena pesada (O_H) y los promotores de las dos hebras (P_I y P_H). También existen otros trectos o pares de bases puntuales que no están directamente implicados en la síntesis de proteínas o de RNAs, como la adenina (A) en la posición 7517 y las 5 pb situadas entre las posiciones 5580 y 5586 (entre los genes tRNA^{Trp} y tRNA^{Ala}) (Chinnery, 2006).

Las dos cadenas de mtDNA tienen una composición asimétrica de las bases, por lo cual se distinguen la cadena pesada (en inglés *Heavy*, H), rica en guaninas (G), y la cadena ligera (en inglés *Light*, L), rica en citosinas (C). Por convención, las posiciones del mtDNA humano se numeran de la 1 a la 16569, tomando como referencia la cadena ligera del primer genoma mitocondrial secuenciado (Anderson et al., 1981), denominado CRS (*Cambridge Reference Sequence*) o secuencia de Anderson;

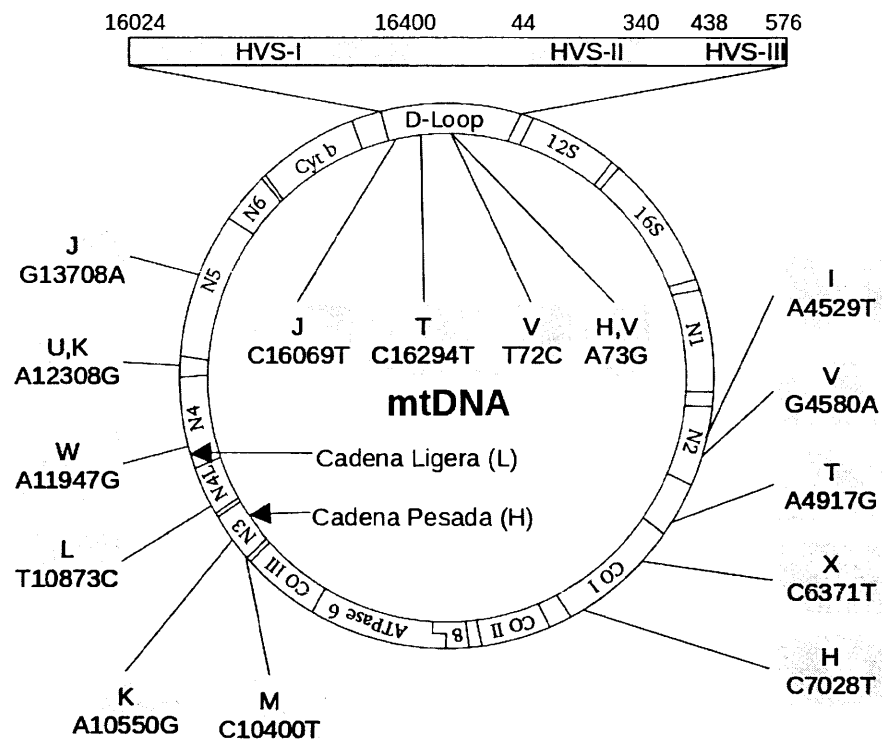
posteriormente revisada por Andrews y colaboradores (Andrews et al., 1999) (rCRS, *revised CRS*).

Los polimorfismos del mtDNA, que incluyen cambios de una sola base (SNPs, en inglés *Single Nucleotide Polimorphisms*) e inserciones/deleciones (indels), se señalan como diferencias (mutaciones) respecto a la rCRS. La tasa de mutación no es constante a lo largo de la molécula de mtDNA, puesto que es considerablemente mayor en las 1,1 kb (kilo bases) de la región control (bases 16024-576) que en las 15,5 kb de la región codificante (bases 577-16023). Estas diferencias se deben a que la región control, aunque juegue un papel importante en la replicación, no es codificante, por lo que puede acumular más mutaciones sin que éstas resulten perjudiciales. Existen también tasas de mutación diferenciales entre posiciones nucleotídicas dentro de la región control (Wakeley, 1993). Algunas posiciones, denominadas puntos calientes o *hot spots*, son hipermutables (por ejemplo las posiciones 146, 150, 152, 195, 16189, 16311, 16362, 16519), mientras que otras son relativamente estables (por ejemplo las posiciones 477, 493, 16108, 16219). También en la región codificante existen posiciones más propensas que otras a la mutación (por ejemplo las posiciones 709, 1719, 3010, 5460, 10398, 11914, 13105, 13708, 15884), causando homoplasias (recurrencias) en la filogenia del mtDNA (van Oven y Kayser, 2009).

La región control incluye tres fragmentos llamados secuencias o regiones hipervariables: HVS-I, HVS-II y HVS-III (en inglés, *Hypervariable Sequences*, HVS), o HVR1, HVR2 y HVR3 (en inglés, *Hypervariable Regions*, HVR) (Brandstätter et al., 2004). El rango de cada una de estas regiones varía ligeramente entre la comunidad científica del ámbito forense (HVS-I: 16024–16365, HVS-II: 73–340 y HVS-III: 438–576) y los estudios de genética de poblaciones, que incluyen rangos más amplios (HVS-I: 16024–16400, HVS-II: 44–340, y HVS-III: 438–576) (Chinnery, 2006). De las tres regiones, la HVS-I es la más variable, con una tasa de mutación dos veces mayor que la HVS-II (Meyer et al., 1999). Últimamente se viene realizando la amplificación del mtDNA completo o de una amplia batería de SNPs con fines identificativos y poblacionales, siendo ambas estrategias mucho más informativas que las regiones hipervariables (Quintáns et al., 2004; Soares et al., 2010; Schonberg et al., 2011). Estos enfoques

resultan prometedores también en el campo del aDNA, especialmente en lo que concierne a la obtención de secuencias mitocondriales completas mediante el empleo de nuevas técnicas de secuenciación masiva disponibles (Knapp y Hofreiter, 2010; Stoneking y Krause, 2011).

Figura I 1 Esquema de la molécula de mtDNA. Se detallan la localización del D-Loop y, en su interior, las regiones hipervariables. También se representa la localización de los principales genes a lo largo de la molécula y las mutaciones asociadas a los principales haplogrupos.



I.2.1.1 CARACTERÍSTICAS DEL ADN MITOCONDRIAL

El mtDNA es una potente herramienta de investigación de las poblaciones humanas debido a sus peculiares características. Entre ellas destacan el elevado número de copias, la heteroplasmia, la ausencia de recombinación, la herencia materna y la elevada tasa de mutación. A continuación se explica brevemente cada una de estas características.

- *Elevado número de copias*

Las células somáticas humanas contienen dos copias de ADN nuclear, heredadas de los progenitores, pero entre centenares y millares de copias de mtDNA (Robin y Wong, 1988). Por ello esta molécula es más fácil de analizar en casos de escasez de muestra o degradación, como es el caso de la genética forense o del ADN antiguo.

- *Heteroplasmia*

Teniendo en cuenta la cantidad de moléculas de mtDNA presente en un organismo humano, es posible que no todas las copias sean idénticas. Este fenómeno se conoce como heteroplasmia y se ha estimado que alrededor del 14 % de la población humana presenta dos variantes de mtDNA con una frecuencia de al menos un 1 % (Tully et al., 2000). Estos datos apuntan a la presencia de un mecanismo muy eficiente de selección de moléculas mitocondriales en los primeros estadios de la oogénesis (Pakendorf y Stoneking, 2005).

- *Herencia materna*

La herencia uniparental de este marcador genético, junto con la consecuente falta de recombinación, incluye una serie de ventajas respecto a los marcadores nucleares autosómicos, tales como el estudio de linajes y ancestros poblacionales. Durante muchos años la herencia exclusivamente materna del mtDNA ha representado un verdadero dogma (Stoneking et al., 1992). Este dogma ha sido revisado desde que se detectó un caso de herencia mitocondrial prevalentemente paterna (Schwartz y Vissing, 2002), poniendo en discusión las inferencias poblacionales basadas en la herencia materna del mtDNA. Sin embargo, se trata del único caso descrito en la literatura y representa una rarísima excepción a la regla. Numerosos estudios sobre la herencia mitocondrial no han vuelto a evidenciar la presencia de herencia paterna, por lo tanto la herencia materna del mtDNA puede seguir considerándose como una constante (Schwartz y Vissing, 2003). Además, se ha demostrado que el mtDNA paterno es selectivamente destruido por el oocito (Manfredi et al., 1997), por lo que su persistencia representa un suceso muy improbable.

- Elevada tasa de mutación

En su conjunto, el mtDNA es más sensible al daño oxidativo debido a su localización fuera del núcleo. Además, carece de la cobertura de proteínas como las histonas, que protegen el ADN nuclear. La replicación de esta molécula es llevada a cabo por una polimerasa mitocondrial propia, que presenta una actividad correctora limitada. Todos los factores citados influyen sobre su tasa de mutación, que es unas 10 veces mayor que la del ADN nuclear. La tasa de mutación en las diferentes regiones del mtDNA no es homogénea, en función de si se trata o no de regiones codificantes. Mientras que la región codificante presenta una tasa de mutación de $1,7 \times 10^{-8}$ sustituciones por sitio y año (Ingman et al., 2000), la de la región control es aún mayor, estimada entre $7,5 \times 10^{-8}$ (Bandelt et al., 2006) y $3,02 \times 10^{-8}$ (Endicott y Ho, 2008) por sitio y año. Sin embargo, como ya se ha comentado, ese valor varía mucho dependiendo de la posición concreta y de los métodos de estimación. De hecho, a nivel de los *hot spots* la tasa de mutación puede alcanzar valores entre 4 ó 5 veces mayores respecto a la media de la región control (Stoneking, 2000; Malyarchuk et al., 2002).

1.2.1.2 VARIABILIDAD GENÉTICA DEL ADN MITOCONDRIAL

Mediante la comparación del mtDNA con la secuencia de referencia (rCRS) se obtiene el haplotipo mitocondrial. Los haplotipos pueden presentar una distribución geográfica o étnica específica y por ello esta información resulta muy interesante en estudios poblacionales. Los haplotipos se agrupan según sus relaciones filogenéticas en haplogrupos, que se designan con letras mayúsculas del alfabeto. Se considera que las secuencias integradas en un mismo haplogrupo mitocondrial proceden de una única molécula ancestral, de la que se han diversificado por mutación.

Los primeros haplogrupos se identificaron en un estudio de mtDNA de americanos nativos, y se llamaron A, B, C y D (Torroni et al., 1993). Así pues los haplogrupos detectados sucesivamente se designaron, por orden, con las restantes letras del alfabeto. En la mayoría de los casos, la asignación de haplogrupos no se puede basar exclusivamente en los datos proporcionados por la región control (haplotipos), aunque

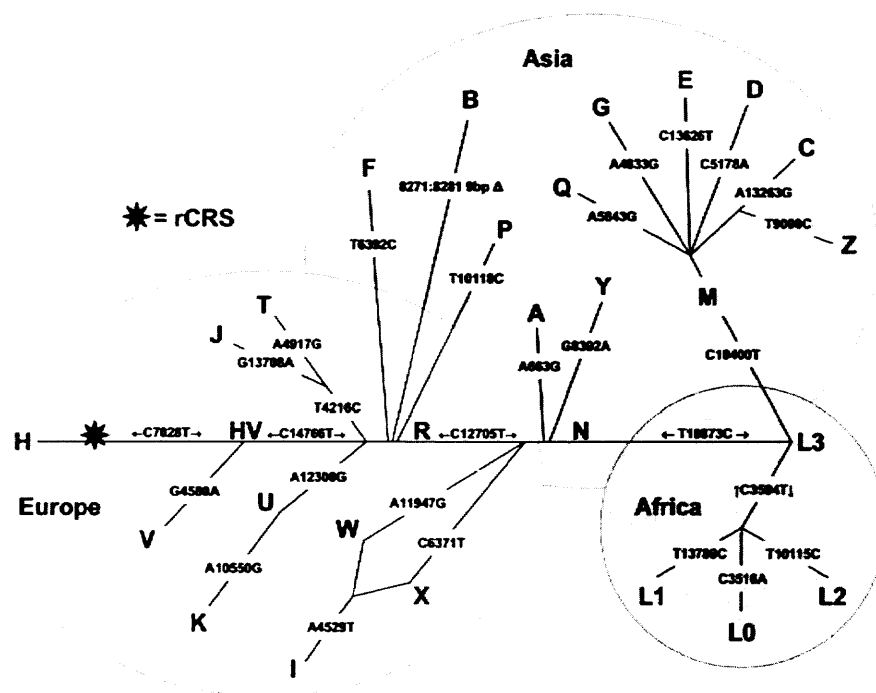
existen haplogrupos por los cuales la asignación basada en la mera información haplotípica puede resultar fiable (por ejemplo los haplogrupos J1b1, K1a9) (van Oven y Kayser, 2009). Por tanto resulta relevante la confirmación del haplogrupo, pronosticado por el haplotipo, mediante SNPs informativos localizados en la región codificante, aunque la identificación de estos SNPs depende de la información filogenética disponible (van Oven y Kayser, 2009). Los haplogrupos también tienen una distribución étnica y geográfica característica. Por lo que respecta a Europa, los principales haplogrupos mitocondriales que se encuentran son: H, J, K, N1, T, U4, U5, V, X y W (Loogväli et al., 2004) (ver **Figura I 2**). Cada haplogrupo puede incluir diferentes sub-haplogrupos (designados por la letra del haplogrupo seguida de un número), que a su vez pueden incluir varios sub-linajes (nombrados por la letra del haplogrupo seguida de un número y una letra). Por ejemplo, dentro del haplogrupo H se han identificado hasta 39 sub-*clusters* que van de H1 a H39; el H2 está a su vez subdividido en H2a y H2b; el primero va desde H2a1 hasta H2a5 (van Oven y Kayser, 2009). Esta compleja nomenclatura que alterna letras y números se debe a su estabilidad en el caso de sucesivas modificaciones en la filogenia (Jobling et al., 2003). En la **Figura I 2** están representados los principales haplogrupos dentro de cada continente.

Muchos de estos haplogrupos presentan una distribución uniforme en Europa, como el haplogrupo H, que es el más frecuente en casi todas las poblaciones europeas (40-60 %) (Richards et al., 2000). Por otro lado existen otros haplogrupos con una distribución muy característica, como el haplogrupo J, muy frecuente en Próximo Oriente, con cierto gradiente de frecuencias decreciente hacia el noroeste de Europa (Richards et al., 2000). Otros haplogrupos, como el V, y ciertos sub-haplogrupos, como el H1, H3 y U5b, presentan picos de frecuencia en la franja cantábrica de la Península Ibérica y en el Norte de Europa (Torroni et al., 2006).

La frecuencia y distribución de los haplogrupos europeos se ha asociado con los principales eventos migratorios de nuestro continente, como se explicará más adelante (ver **Apartado I.4.1.2**). Por esta razón y gracias a su abundancia en la célula, el mtDNA es ampliamente utilizado en estudios de genética de poblaciones antiguas.

En los últimos 20 años, se han publicado estudios poblacionales humanos que se remontan hasta el mesolítico. A pesar del reducido número de muestras estudiadas, de la limitada disponibilidad de restos antiguos y de las dificultades asociadas a su análisis, los resultados obtenidos hasta la fecha han contribuido al conocimiento de la composición genética de nuestros antepasados (ver **Apartado I.4.1.3**).

Figura I 2 Relación filogenética entre los principales haplogrupos mitocondriales en correspondencia de las regiones geográficas (imagen modificada a partir de <http://mitomap.org>).



Además, el mtDNA permite la reconstrucción de genealogías y relaciones familiares a lo largo de la línea materna. Por ello, su análisis puede aplicarse para examinar de manera prospectiva las relaciones existentes en casos de enterramientos múltiples. El estudio del mtDNA permite excluir la hipótesis de maternidad si se presentan haplotipos diferentes. En la situación opuesta, en función del haplotipo encontrado, es posible calcular la probabilidad de maternidad pero con una probabilidad reducida (ver **Apartado I.4.2**).

I.2.2 El ADN nuclear

El ADN nuclear (nuDNA) incluye la mayoría de la información genética para las funciones vitales de los individuos. Su genoma está organizado en 22 parejas de cromosomas autosómicos y una pareja de cromosomas sexuales, XX en mujeres y XY en varones. En total suman unas 3.000 Mb (Mega bases o millones de bases), que contienen alrededor de 25.000 genes.

En función de las aplicaciones concretas, se suelen estudiar diferentes regiones polimórficas del genoma. Para estudiar relaciones familiares cercanas o identificar un individuo, se analizan marcadores altamente variables repetidos en tándem, como los microsatélites o STRs (en inglés, *Short Tandem Repeats*), cuya tasa de mutación es de alrededor de 10^{-3} - 10^{-4} por locus y por generación (Zhivotovsky et al., 2004). Otros marcadores muy empleados son los SNPs, que consisten en mutaciones puntuales a nivel de una base concreta y que tienen una tasa de mutación de alrededor de 10^{-7} - 10^{-8} por locus y generación (Jobling et al., 2003). Se han identificado más de 10 millones de SNPs en el genoma humano, representando la mayor fuente de variabilidad genética, con aplicaciones en diferentes ámbitos, como el campo forense, la genética de poblaciones, el estudio y diagnóstico de enfermedades y sobre todo el estudio de asociaciones con enfermedades comunes y caracteres complejos (Beckmann et al., 2007). Sin embargo, más recientemente, se ha demostrado que existe otra fuente responsable por lo menos del 10 % de la variabilidad genómica (Redon et al., 2006; Wong et al., 2007): los CNVs (en inglés, *Copy Number Variants*), que incluyen modificaciones estructurales de grandes fragmentos del genoma (Beckmann et al., 2007). Se trata de variantes menos frecuentes que los SNPs, pero que implican cambios estructurales importantes, desde 1 Kb hasta varias Mb (Redon et al., 2006; Wong et al., 2007). Así los CNVs se están convirtiendo en una herramienta cada vez más empleada en el estudio de características fenotípicas, como los caracteres complejos y las enfermedades genéticas (Beckmann et al., 2007).

En general, los trabajos relativos al estudio de marcadores nucleares en el campo del aDNA se están desarrollando cada vez más gracias a las técnicas especialmente

enfocadas a la recuperación de ADN degradado y a los avances en genética molecular. Además, se han puesto a punto nuevas técnicas de secuenciación masiva, denominadas genéricamente Secuenciación de Nueva Generación (*Next Generation Sequencing*, NGS) que permiten analizar potencialmente el genoma completo de un organismo de una forma rápida (Knapp y Hofreiter, 2010). Dichas técnicas permiten obtener gran cantidad de información genética a partir de muy bajas concentraciones de ADN original. Sin embargo, el análisis de los resultados suele complicarse a la hora de separar el ADN endógeno del contaminante, especialmente en el caso del ADN humano, y de identificar las lesiones *post-mortem* (Millar et al., 2008). Aunque sus aplicaciones en el campo del aDNA están todavía en una fase inicial, ya se han publicado algunos trabajos de gran relevancia, como por ejemplo el estudio de gran parte del genoma del hombre de Neandertal (Green et al., 2010) y del mamut (Poinar et al., 2006). Mediante técnicas similares, también ha sido posible recuperar una proporción muy elevada del genoma procedente de un individuo siberiano de hace 4.000 años (Rasmussen et al., 2010) y de un hombre arcaico de las mismas regiones, el homínido de Denisova (Krause et al., 2010).

I.2.2.1 MARCADORES AUTOSÓMICOS

Los marcadores autosómicos se sitúan en las 22 parejas de cromosomas heredados de los progenitores. Debido al proceso de recombinación al que están sometidos entre generaciones, resulta muy complicado reconstruir el pasado de las poblaciones mediante una filogenia, como en el caso de los marcadores uniparentales (del mtDNA y el cromosoma Y).

En general, los marcadores autosómicos más empleados en el campo del aDNA se pueden agrupar en dos categorías: aquellos relacionados con caracteres fenotípicos (por ejemplo, con enfermedades) y los marcadores neutros, que se emplean con una finalidad identificativa. En el primer caso, uno de los marcadores más estudiados en especímenes humanos antiguos es el gen de la lactasa, cuya difusión está relacionada con una ventaja evolutiva en poblaciones productoras de productos lácteos (Burger et al., 2007; Malmström et al., 2010; Lacan et al., 2011; Nagy et al., 2011). En el segundo

caso, estos marcadores hacen posible la identificación de relaciones familiares cercanas, como la paternidad, la maternidad o la hermandad, proporcionando probabilidades elevadas. Se trata de los mismos sistemas empleados en el campo forense, que actualmente es posible analizar gracias a *kits* especialmente diseñados para muestras muy degradadas (ver **Apartado I.4.2**).

I.2.2.1.1 EL CROMOSOMA Y

El cromosoma Y representa tan solo el 2-3 % del genoma humano haploide y posee muy pocos genes codificantes: los de la fertilidad masculina (DAZ y RBMY), el gen SRY (en inglés *Sex-determining Region of Y*) y otros situados en la PAR (región pseudoautosómica), la única que presenta un homólogo en el cromosoma X. El cromosoma Y es en esencia un marcador de linaje, puesto que toda su porción no recombinante (NRY, en inglés *Non Recombining region of the Y chromosome*) se transmite de forma invariable a lo largo de la línea paterna, variando exclusivamente a través del acúmulo progresivo de mutaciones. Como en el caso del mtDNA, también se habla de haplotipos del cromosoma Y, que en este caso se basan en la información proporcionada por una batería de STRs. Asimismo, para el análisis filogenético de este marcador, es importante completar esta información con aquella proporcionada por sistemas con una tasa de mutación menos elevada, los SNPs (Jobling y Tyler-Smith, 2003).

I.2.2.1.2 VARIABILIDAD GENÉTICA DEL CROMOSOMA Y

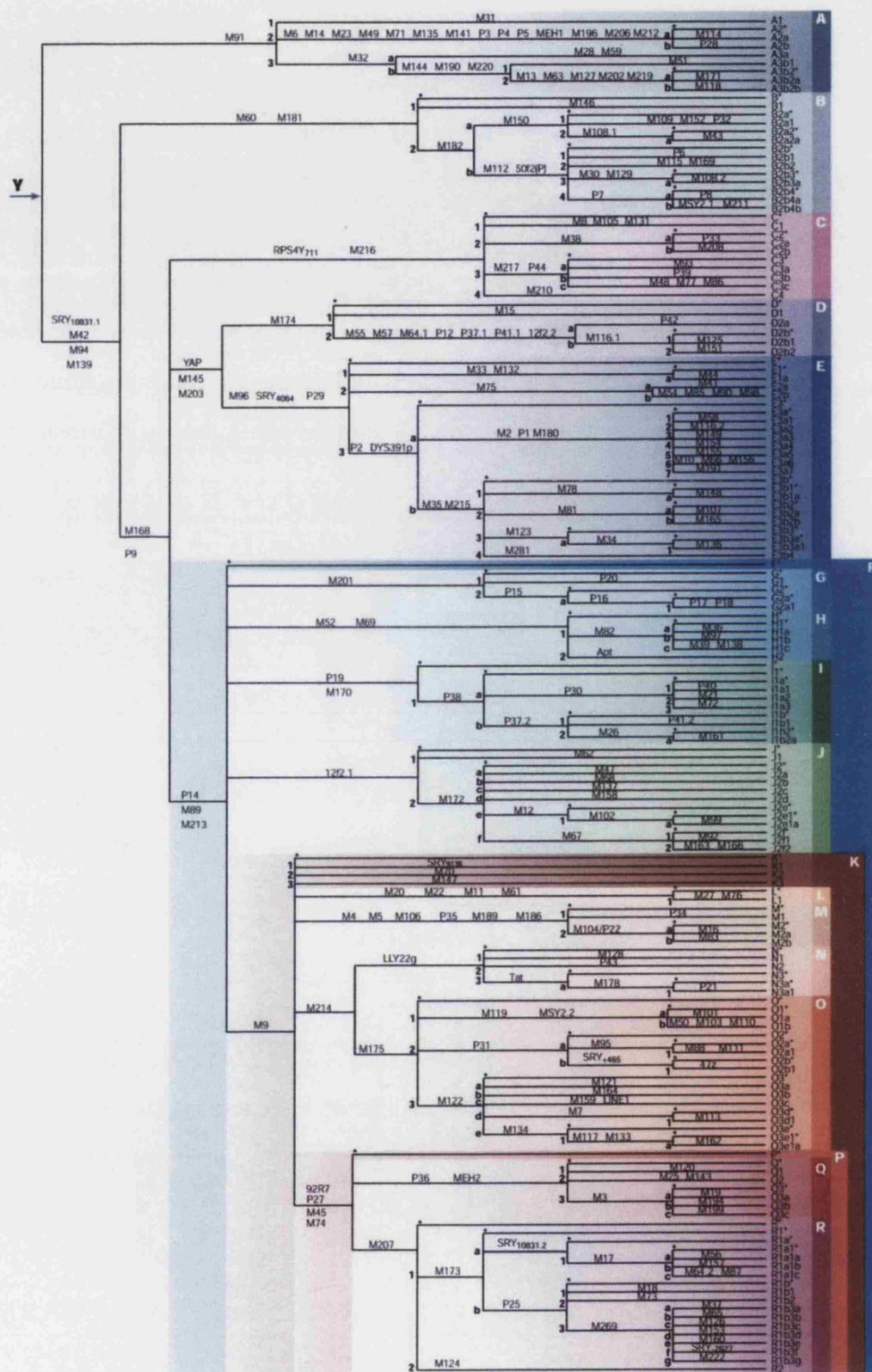
Para el cromosoma Y, el esfuerzo conjunto de diferentes grupos de investigación del consorcio YCC (*Y Chromosome Consortium*, 2002, <http://ycc.biosci.arizona.edu>) permitió definir el árbol filogenético de este cromosoma, basado en marcadores bialélicos (SNPs). En concreto, se seleccionaron aquellos SNPs más estables, denominados UEPs (en inglés, *Unique Event Polymorphisms*). Gracias al empleo de

estos sistemas fue posible definir los principales haplogrupos del cromosoma Y, representados en la **Figura I 3** (Jobling y Tyler-Smith, 2003).

La nomenclatura del Cromosoma Y incluye haplogrupos que van desde el A hasta el T. En este caso, a diferencia del mtDNA, el orden alfabético refleja la profundidad filogenética, siendo el haplogrupo A el antecesor común de los demás haplogrupos. Sin embargo, los linajes más antiguos (A y B) están muy poco representados en poblaciones africanas actuales, donde el haplogrupo más frecuente es el U, que se ha asociado tradicionalmente a la expansión Bantú durante la Edad del Hierro (Jobling y Tyler-Smith, 2003).

En lo que concierne a Europa, los principales haplogrupos son el R1b, el R1a, el I, el N, el E1b1b1 y el J. El primero de ellos (R1b) es el más común de todo el continente, especialmente en la parte occidental (Semino et al., 2000; Balaresque et al., 2010; Myres et al., 2011). Su frecuencia es mayor en las áreas atlánticas y disminuye hacia Europa central y oriental, donde el R1a presenta frecuencias mayores (Semino et al., 2000; Passarino et al., 2001; Peričić et al., 2005). El haplogrupo N es común solo en el noreste europeo mientras que ciertos sub-haplogrupos del E1b1b1 caracterizan a regiones específicas, como la Península Ibérica y Kosovo. El haplogrupo I presenta diferentes focos de frecuencias elevadas (Rootsi et al., 2004), mientras que el J presenta cierta clina sureste noroeste (Semino et al., 2004).

Figura 1 3 Árbol filogenético de los haplogrupos del cromosoma Y (imagen de Jobling y Tyler-Smith, 2003).



En el campo de la genética de poblaciones actuales se han descrito las distribuciones y dataciones moleculares de los haplogrupos del cromosoma Y. Estos datos han contribuido al conocimiento del pasado de las poblaciones humanas desde el punto de vista de los linajes masculinos. En lo que respecta a las aplicaciones de este marcador en el campo del aDNA, éstas están limitadas por su escasa abundancia en la célula y por la falta de *kits* comerciales de amplificación especialmente diseñados para muestras muy degradadas. El cromosoma Y representa un marcador de gran interés en este campo, puesto que permite sondear procesos migratorios desde la perspectiva de linajes paternos. Por otra parte, dado que el cromosoma Y incluye también marcadores con elevada tasa mutacional (los STRs), su estudio permite establecer relaciones de parentesco a lo largo de la línea paterna asociadas a probabilidades mucho más elevadas que en el caso del mtDNA.

1.2.2.2 EL CROMOSOMA X

El cromosoma X presenta un mayor tamaño respecto al cromosoma Y y contiene unas 153 Mb, representando alrededor del 5 % del genoma. Aunque se presente en doble copia en las mujeres, existe un mecanismo de compensación por el que solamente uno de los dos cromosomas queda activado. Este cromosoma contiene alrededor de unos 1.300 genes, una cantidad muchísimo mayor respecto al cromosoma Y, pero bastante reducida comparada con otros cromosomas nucleares.

La principal diferencia entre los cromosomas sexuales es que el Y carece de recombinación durante la meiosis, si se excluyen las pequeñas zonas de recombinación con el X (PAR) (Quintana-Murci y Fellous, 2001). En el caso de las mujeres, el cromosoma X recombina con su homólogo, comportándose como todas las demás parejas de cromosomas nucleares. Sin embargo, la diversidad genética del cromosoma X es menor respecto a los cromosomas autosómicos, dado que su población efectiva (N_e) es tres cuartas partes la de los autosomas, en cuanto se presenta en copia única en los individuos masculinos.

A pesar de la presencia de recombinación, el cromosoma X contiene regiones caracterizadas por un elevado desequilibrio de ligamiento (LD, en inglés *Linkage Disequilibrium*) y a partir de cada una de ellas sería posible construir filogenias independientes (Gabriel et al., 2002; Schaffner, 2004). La presencia de recombinación podría representar también una ventaja: el cromosoma X puede aportar información acerca de centenares o incluso millares de fragmentos independientes heredados en bloque, mientras que el mtDNA y el cromosoma Y solamente de uno (Schaffner, 2004). Esto representa el principal potencial de este cromosoma en genética de poblaciones, con aplicaciones que van desde problemáticas globales, como el origen del hombre moderno, hasta la historia humana más reciente y local, como el estudio del poblamiento de América. Sin embargo, su empleo en este tipo de estudios es muy reducido debido a las dificultades asociadas a la construcción de diferentes filogenias (Tomas et al., 2008; Pereira et al., 2011). Por ello, la mayoría de este tipo de investigaciones se basa en marcadores uniparentales, el mtDNA y el cromosoma Y, cuya filogenia ha sido ampliamente estudiada.

El potencial del cromosoma X es elevado, también porque representaría la historia de la población humana y no solamente de uno de los sexos. Gracias al desarrollo de las técnicas de secuenciación masiva, el interés y el empleo de este marcador en genética de poblaciones podría ser favorecido (Schaffner, 2004). Actualmente está también ampliamente empleado en genética forense, porque debido a su patrón de herencia permite estudiar relaciones familiares en las que la genealogía no esté completa.

La principal dificultad asociada al análisis del cromosoma X en aDNA reside en su baja concentración comparada con el mtDNA, como ocurre en el caso del cromosoma Y. Sin embargo, gracias al desarrollo de nuevas técnicas y protocolos, su estudio en poblaciones antiguas presenta un gran potencial.

I.3 EL ADN ANTIGUO: HISTORIA Y PROBLEMÁTICAS ASOCIADAS A SU ANÁLISIS

I.3.1 Historia del ADN antiguo

A mediados de los años 80 se publicaron los primeros dos trabajos de ADN antiguo, que se basaban en técnicas de clonación bacteriana para la amplificación de pequeños fragmentos de mtDNA de tejidos momificados (Higuchi et al., 1984; Pääbo, 1985). Sin embargo, las limitaciones de dichas técnicas, la elevada tasa de contaminación por ADN exógeno y la escasa cantidad y calidad de ADN endógeno, no permitieron la replicación de los experimentos, criterio que fundamenta y valida cualquier disciplina científica (Handt et al., 1994).

En la segunda mitad de la misma década, se desarrolló la técnica de amplificación *in vitro* mediante la reacción en cadena de la polimerasa o PCR, en inglés *Polymerase Chain Reaction* (Mullis y Faloona, 1987), que supuso un avance tecnológico fundamental en el campo de la biología molecular. Además facilitó enormemente el consiguiente desarrollo de estudios de aDNA, puesto que hacía posible la amplificación y análisis de ADN a partir de cantidades muy pequeñas de moléculas degradadas. Así pues la amplificación *in vitro* fue inmediatamente aplicada a muestras antiguas. Entre los primeros trabajos de aDNA amplificados por PCR, destaca el de Pääbo y Wilson (Pääbo y Wilson, 1988) en el que se hace hincapié en las ventajas de esta nueva técnica. Estos investigadores consiguieron replicar el experimento de Higuchi y colaboradores (Higuchi et al., 1984) detectando sustituciones que no se habían podido identificar mediante la clonación bacteriana, debido a que habían sido corregidas por los mecanismos de reparación de las bacterias.

Estos primeros resultados fueron acogidos con entusiasmo por la comunidad científica y diferentes grupos de investigación emprendieron nuevos estudios sobre muestras antiguas. En 1989 se demostró la posibilidad de recuperar material genético a partir de

huesos y dientes de hasta 5.000 años de antigüedad (Hagelberg et al., 1989). Dichos materiales representan el tipo de muestra más abundante en el registro arqueológico, dado que se trata del material que mejor se conserva en los vertebrados, ampliando así las posibilidades de aplicaciones de la genética al mundo de la arqueología y a la genética forense.

Sin embargo, la gran sensibilidad de la PCR conlleva un elevado riesgo de contaminación con ADN exógeno de los propios manipuladores o procedente de moléculas previamente amplificadas en el laboratorio. Por ello, muchos de los primeros trabajos publicados fueron posteriormente desacreditados, demostrándose que no era posible su replicación y que se trataba de contaminaciones con ADN actual de diferente origen. Estos casos afectaron a publicaciones de secuencias de millones de años de antigüedad que habían suscitado mucho entusiasmo y polémica. Golenberg y colaboradores (Golenberg et al., 1990) amplificaron ADN procedente de una hoja de magnolia de 17-20 millones de años, cuyo análisis filogenético relacionaba las secuencias antiguas con las especies de *Magnolia* actuales, apoyando la autenticidad de los resultados. Sin embargo, este estudio fue llevado a cabo sin ningún control de posibles contaminaciones. Además no fue posible replicar los resultados de forma independiente y, por estas razones, fue ampliamente criticado (Pääbo y Wilson, 1991; Lindahl, 1993).

Una gran variedad de restos incluidos en ámbar fueron analizados porque se creía que esta resina permitía una desecación rápida del organismo e inhibía toda actividad bacteriana, favoreciendo la preservación del ADN (Cano et al., 1993; Poinar, 1994). Así pues se obtuvieron secuencias de ADN a partir de restos de abejas, termitas y plantas (DeSalle et al., 1992; Cano et al., 1993; Poinar et al., 1993). Incluso se creyó haber identificado y analizado bacterias fósiles (Cano et al., 1994), aunque en realidad se trataba de contaminaciones (Yousten y Rippere, 1997). Ulteriores estudios demostraron que la conservación del ADN durante millones de años es altamente improbable (Stankiewicz et al., 1998).

El trabajo con más impacto mediático probablemente fue la recuperación de mtDNA a partir de fósiles de dinosaurios de hace 80 millones de años (Woodward et al., 1994).

Se confió en la autenticidad de las secuencias puesto que nunca se habían descrito antes ni en mamíferos ni en reptiles. Posteriormente se demostró que se trataba de una contaminación por ADN humano actual, procedente de una porción del ADN mitocondrial integrada en el genoma nuclear y que presentan ciertos individuos (NUMTs, en inglés *nuclear mitochondrial DNA*) (Zischler et al., 1995). Desde 1995 no se han vuelto a publicar resultados de muestras con millones de años de antigüedad, dado que se ha demostrado que el límite máximo para la preservación y amplificación del aDNA ronda los 100.000 años (Lindahl, 1993, 1997).

El equipo de Svante Pääbo puso de manifiesto los inconvenientes asociados a los estudios de aDNA publicados, destacando las limitaciones de las técnicas empleadas y de las instalaciones, así como la falta de control de las contaminaciones (Pääbo, 1989; Pääbo et al., 1989). Hay que tener en cuenta que una molécula de ADN fresco procedente de una fuente contaminante tiene más probabilidades de ser amplificada que las moléculas antiguas endógenas, que están degradadas y dañadas. Además existe la posibilidad de obtener secuencias mixtas o quimeras de aDNA, resultado de la co-amplificación de moléculas endógenas y contaminantes, mediante lo que se conoce como *Jumping PCR* (ver **Apartado I.3.3.2**) (Pääbo et al., 1989). Existe el riesgo de no detectar este fenómeno, puesto que estas quimeras podrían tener cierto sentido filogenético. Por otra parte, el ADN dañado puede provocar la inserción de adeninas durante la PCR así como errores de incorporación en motivos repetitivos (DeSalle et al., 1993).

Hay que tener en cuenta que las fuentes de contaminación no incluyen solamente la fase de análisis en el laboratorio (personal del laboratorio, reactivos, moléculas amplificadas flotantes), sino también a todas aquellas personas y sustancias que hayan entrado en contacto con las muestras desde la excavación hasta su llegada al laboratorio (Handt et al., 1994). La dificultad en el control de posibles contaminaciones en el análisis de restos antiguos es aún mayor en el caso de especímenes humanos, así que la primera prueba que testifica la autenticidad de un resultado es la replicación del mismo de forma independiente.

Entre finales de los 90 y principios del 2000 se comenzó a hacer hincapié en la importancia de ciertos criterios en la validación de los resultados obtenidos a partir de muestras antiguas, los cuales fueron asimilados por las revistas de impacto como requisitos para la publicación de estudios de aDNA. Uno de los trabajos más importantes en el establecimiento y empleo de estos criterios es el de Krings y colaboradores (Krings et al., 1997), en el que se demostró la autenticidad de los resultados obtenidos a partir del primer espécimen de Neandertal analizado. Más recientemente se han publicado diferentes revisiones exhaustivas de todos los criterios necesarios para la validación de resultados (Poinar et al., 1993; Cooper y Poinar, 2000; Pääbo et al., 2004) (ver **Apartado I.3.3.3**).

A pesar de sus comienzos controvertidos, el ADN antiguo se puede considerar una disciplina científica establecida: se han mejorado muchísimo los estándares, así como el conocimiento de la naturaleza bioquímica de los procesos diagenéticos y del daño molecular del ADN (Willerslev y Cooper, 2005). Se han publicado además estudios rigurosos que contribuyen al conocimiento de procesos evolutivos, proporcionando información de primera mano para testar modelos e hipótesis en genética de poblaciones, paleoecología y evolución (Willerslev y Cooper, 2005).

I.3.2 Factores de degradación del ADN antiguo

La recuperación de ADN a partir de muestras antiguas está sujeta principalmente a la preservación de la muestra. Existen diferentes mecanismos que afectan al material genético de las muestras biológicas antiguas debido a los propios fenómenos de muerte celular y a las condiciones medioambientales y físico-químicas del entorno de deposición a las que ha estado sometida. En las células, los mecanismos enzimáticos de reparación dejan de funcionar poco tiempo después de la muerte del organismo. Para los humanos la autólisis celular empieza a los 4 ó 5 minutos de la muerte (Vass, 2001), cuando la concentración de oxígeno en la sangre disminuye a la vez que aumenta la de dióxido de carbono, baja el pH y se acumulan los productos de

degradación celular. Al mismo tiempo, las enzimas líticas celulares comienzan a degradar la célula desde su interior provocando la rotura de la membrana celular y nuclear, liberando enzimas intracelulares como las nucleasas, que degradan específicamente el material genético. Tras el proceso de autólisis se inicia la colonización del tejido muerto por microorganismos, hongos e insectos, que completan el proceso degradativo (Gilbert, 2006).

Todas estas modificaciones celulares conllevan la degradación del material genético, cuyas modificaciones se denominan diagenéticas o *post-mortem*. El proceso de degradación puede ser muy rápido con una disminución importante de ADN amplificable en tan solo pocas semanas o meses. Existen diferentes mecanismos que ralentizan la degradación, favoreciendo la preservación de los tejidos y del ADN, como una rápida desecación (momificación natural), una elevada concentración de sales o bajas temperaturas (Hofreiter et al., 2001 b). Estos factores inhiben o ralentizan la actividad enzimática de las endonucleasas.

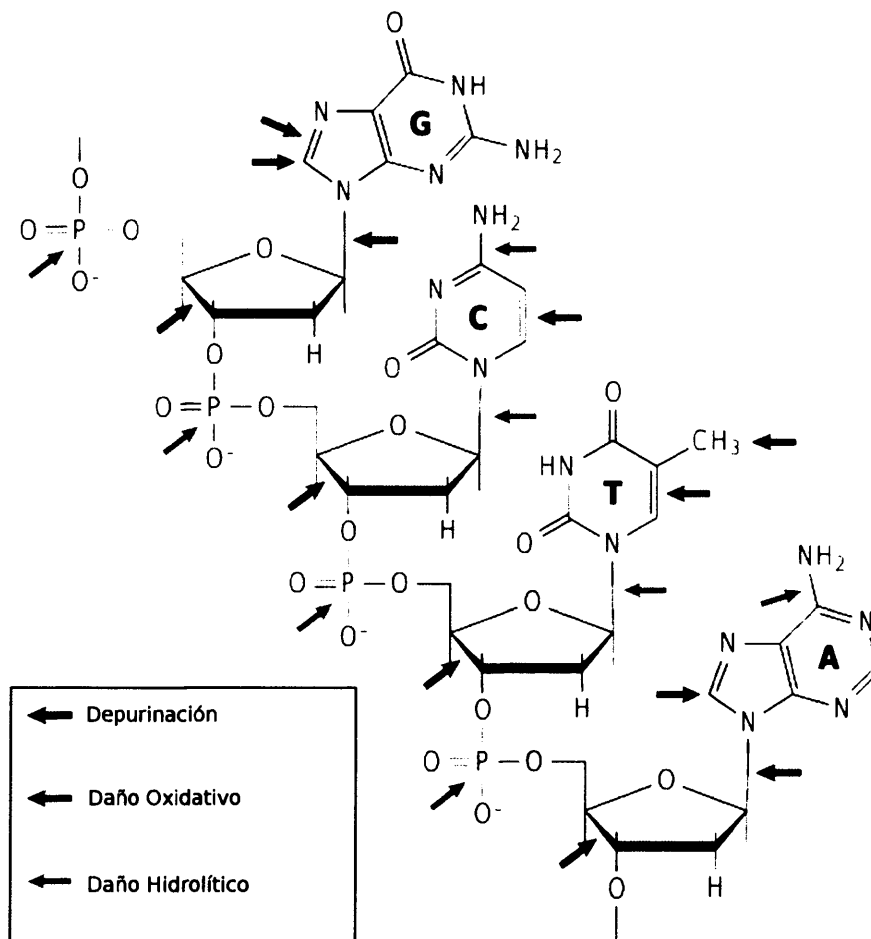
Existen muchos mecanismos que actúan sobre la molécula de ADN provocando fragmentación y modificaciones estructurales y químicas que pueden incluso impedir su análisis. A continuación se describen los principales procesos diagenéticos responsables de la degradación del ADN.

I.3.2.1 DEGRADACIÓN ENZIMÁTICA E HIDRÓLISIS

Los mecanismos hidrolíticos y de degradación enzimática son los principales responsables de la fragmentación de las hebras de ADN, reduciendo su longitud a unas 100-200 pb (Pääbo et al., 2004). En primer lugar, la rotura de ciertos orgánulos celulares, como los lisosomas, libera enzimas autolíticas como las *endo* y las *exonucleasas* (Pääbo, 1989). Las primeras cortan las cadenas de ADN en diversos puntos y las segundas eliminan los nucleótidos a partir de los extremos de la cadena. A continuación, el daño hidrolítico actúa sobre los enlaces fosfodiéster y N-glicosídicos provocando respectivamente la rotura de la hebra o la eliminación de una base nitrogenada (Lindahl, 1993) (ver **Figura I 4**). En el segundo caso se habla de

depurinización o *depirimidinación*, en función de si la base eliminada es una purina (adenina o guanina) o una pirimidina (citosina o timina). La hidrólisis afecta principalmente a los enlaces N-glicosídicos de las purinas, siendo la tasa de *depurinización* 20 veces mayor que la de *depirimidinación* (Lindahl, 1993). Después de la eliminación de una base se suceden mecanismos de reorganización estructural que llevan a la rotura de la cadena de ADN o incluso a la pérdida de bases adyacentes.

Figura 1 4 Principales modificaciones *post-mortem* del ADN. Las flechas rojas indican los principales sitios de depurinación, las azules los de daño oxidativo y las verdes los de hidrólisis (imagen modificada a partir de Hofreiter et al. 2001).



Otro punto susceptible al ataque hidrolítico es el enlace glicosídico del grupo amino de la base nitrogenada. Este proceso se denomina desaminación hidrolítica y afecta principalmente al grupo amino de las pirimidinas, siendo menos frecuente en las purinas. Se trata de un proceso que no afecta directamente a la estructura y longitud de la hebra del ADN, como ocurre en el caso de la depurinización, puesto que la estructura de la doble hélice consigue proteger de la desaminación las bases en su interior (Lindahl, 1993). Se ha estimado que la frecuencia de esta reacción en las cadenas sencillas es hasta cien veces más probable que en la doble hélice (Krickler y Drake, 1990).

La desaminación de las bases nitrogenadas provoca la transformación de unos nucleótidos en otros. Por ejemplo, la adenina se transforma en hipoxantina, un análogo de la guanina (Lesión de tipo 1); y la citosina en uracilo, un análogo de la timina (Lesión de tipo 2) (ver **Figura I 5**). Estas modificaciones se denominan *Miscoding Lesions* dado que durante la amplificación del ADN por PCR inducen a la incorporación de nucleótidos incorrectos en la cadena complementaria (Hofreiter et al., 2001 b; Gilbert et al., 2003 b). A la hora de analizar el daño molecular de una secuencia amplificada, es posible detectar cuatro cambios distintos de bases: G→T, T→C, A→G y G→A. Se trata de transiciones entre purinas o pirimidinas causadas por los dos mecanismos de desaminación descritos, y que producen uno u otro cambio dependiendo de la cadena del ADN (directa o reversa) en la que se originen (ver **Figura I 5**).

Sin embargo, recientemente se ha demostrado mediante diferentes métodos que las Lesiones de tipo 2 son las principales, si no las únicas, responsables del daño molecular detectado en secuencias de ADN antiguo (Brotherton et al., 2007; Gilbert et al., 2007). En la **Figura I 6** se detallan los efectos de las *Miscoding Lesions* en la cadena directa o *forward*, utilizada como referencia para la lectura de secuencias de ADN.

Figura I 5 Mecanismos de actuación de la desaminación hidrolítica. Se detallan las lesiones de tipo 1 y de tipo 2.

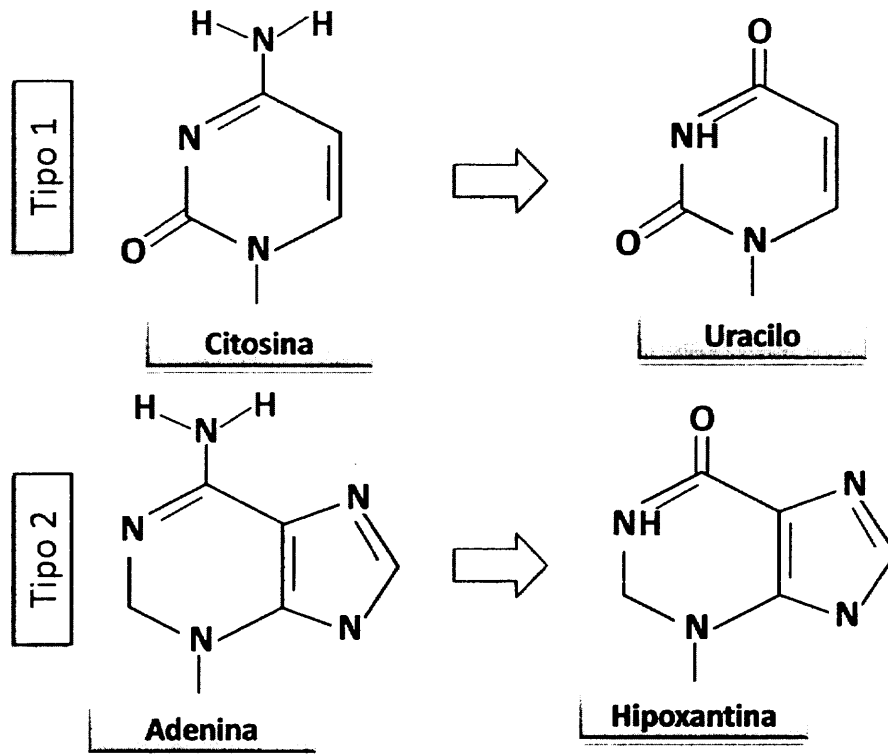
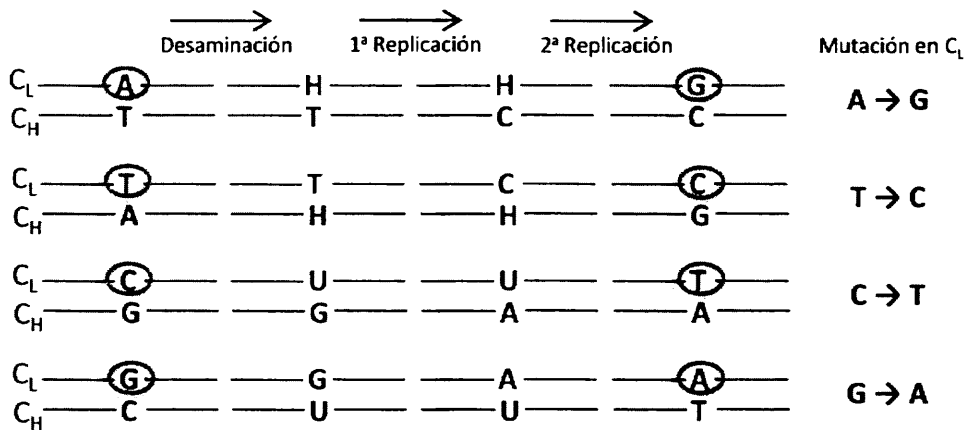


Figura I 6 Mutaciones asociadas a la desaminación hidrolítica de tipo 1 y 2. Se ilustran los distintos resultados dependiendo de si la *Miscoding Lesion* tiene lugar en la cadena ligera (C_L) o en la pesada (C_H).



Como ya se ha comentado anteriormente, existen unas posiciones del mtDNA que presentan una tasa de mutación más elevada y que se denominan puntos calientes mutacionales o *molecular hot spots*. Diferentes autores han estudiado este fenómeno

para comprobar la presencia de un comportamiento similar en el caso del daño molecular. Estos trabajos se fundamentan en la hipótesis de que determinadas posiciones del mtDNA que se encuentran menos protegidas *in vivo* también lo están durante el proceso diagenético (Heyer et al., 2001). Uno de los primeros estudios acerca de este tema afirma que la distribución del daño molecular en el mtDNA de osos de las cavernas no sigue una pauta específica sino que se distribuye al azar (Hofreiter et al., 2001 a). Poco después fue posible identificar estas posiciones en la HVR-I del mtDNA humano (Gilbert et al., 2003 b) y posteriormente también en bisontes (Gilbert et al., 2005). El conocimiento de las posiciones más sujetas al daño molecular puede ayudar en la interpretación de los resultados y a distinguir entre mutaciones endógenas y daño molecular (Gilbert et al., 2003 a).

1.3.2.2 OXIDACIÓN

El ADN puede dañarse también mediante reacciones oxidativas, por la acción de radicales libres que se originan tras la exposición a radiaciones ionizantes, luz ultravioleta y procesos de degradación por parte de bacterias y hongos (Poinar, 2002). Los radicales libres inducen daños estructurales importantes a nivel de los dobles enlaces de las bases nitrogenadas y pueden provocar incluso su rotura (ver **Figura I 4**). También las moléculas de hidrógeno de la desoxirribosa representan una diana del daño oxidativo, con la consecuente rotura del anillo. Otra de las modificaciones provocadas por los radicales libres es la transformación de las bases nitrogenadas, sobre todo de las pirimidinas, en hidantoínas (Höss et al., 1996). Las principales formas de hidantoínas asociadas al daño oxidativo son la 5-OH-5-MeHyd, un derivado de la timina, y la 5-OH-Hyd, un derivado de la citosina (Höss et al., 1996).

El resultado de las modificaciones generadas por oxidación no es la fragmentación de la cadena de ADN sino el bloqueo de la polimerasa durante la PCR, actuando por lo tanto como modificaciones inhibitorias.

I.3.2.3 PUENTES CRUZADOS O "CROSSLINKS"

Además de las modificaciones que implican la rotura de enlaces, existen otras que provocan la formación de nuevos enlaces, denominados puentes cruzados o *crosslinks*. Se han identificado enlaces intramoleculares, entre diferentes cadenas de ADN, e intermoleculares, entre ADN y otras biomoléculas como las proteínas. Estas últimas consisten principalmente en reacciones de condensación entre la desoxirribosa de las bases nitrogenadas y los grupos amino de las proteínas. Los derivados de esta última reacción se identificaron por cromatografía de gases combinada con espectrometría de masas en coprolitos (Poinar et al., 1998) y se conocen como *productos de Maillard* (Lindahl, 1993). Los puentes cruzados también impiden la amplificación del ADN por PCR, bloqueando la polimerasa. Sin embargo, se identificó un compuesto capaz de romper estos enlaces, el bromuro de N-fenilacil-tiazolio o PTB (Vasan et al., 1996), incrementando la probabilidad de obtención de productos de amplificación. El empleo del PTB permitió la recuperación de ADN a partir de coprolitos de unos 20.000 años (Vasan et al., 1996) así como de huesos de Neandertal de más de 40.000 (Krings et al., 2000).

I.3.3 Problemáticas de los estudios de ADN antiguo

I.3.3.1 JUMPING PCR

Como ya se ha mencionado, una consecuencia del daño molecular es la presencia de lesiones que impiden el funcionamiento de la polimerasa, enzima responsable de la PCR. En esos puntos la amplificación se interrumpe, pero podría continuar empleándose como molde otra cadena de ADN diferente. Este mecanismo se denomina *Jumping PCR*, puesto que la polimerasa salta de una cadena molde a otra pudiendo producir secuencias quiméricas, entre diferentes combinaciones de cadenas endógenas, contaminantes y dañadas. El *Jumping PCR* es un fenómeno recurrente en el campo del ADN antiguo. Su identificación se basa en el análisis detallado de los

resultados de clonación bacteriana, así como en la amplificación repetida de diferentes extractos y en la cuantificación de ADN.

I.3.3.2 INHIBIDORES DE LA REACCIÓN DE PCR

Existen diferentes mecanismos y varios factores que pueden disminuir o impedir la amplificación de ADN por PCR. Las moléculas responsables de este fenómeno se denominan inhibidores y suelen actuar bloqueando la actividad enzimática de la polimerasa durante la PCR (Wilson, 1997).

Clásicamente se han identificado dos categorías de inhibidores, aquellos procedentes de entorno de deposición y los resultantes de la propia degradación del organismo. Entre los primeros están los ácidos húmicos y fúlvicos del suelo, que pueden llegar a ser co-purificados junto con el ADN (Pääbo, 1989) y se ha demostrado que tienen un gran poder inhibidor (Tuross, 1994). Entre los segundos se encuentran los residuos de porfirinas. Las porfirinas se pueden encontrar en tejidos animales blandos y restos vegetales, y constituyen una familia de moléculas que derivan de la porfirina, un compuesto tetrapirrólico. Las porfirinas actúan como inhibidores de la PCR puesto que pueden formar complejos mediante la unión a diversos iones, entre los cuales se encuentra el magnesio, necesario para el funcionamiento de la ADN polimerasa (Higuchi et al., 1984). Como ya se ha mencionado anteriormente (**Apartados I.3.2.2 e I.3.2.3**), otros productos de degradación del ADN, como las hidantoínas y los productos de la reacción de *Maillard*, pueden comportarse también como inhibidores de la PCR.

I.3.3.3 CONTAMINACIÓN Y CRITERIOS DE AUTENTICIDAD

Las muestras antiguas contienen ADN fragmentado, dañado y en escasas cantidades, lo que hace que se trate de un material genético de difícil recuperación y muy susceptible a la contaminación por ADN exógeno. Por otra parte, la técnica empleada para la recuperación de información genética a partir de restos antiguos, la PCR, resulta altamente sensible a cantidades mínimas de ADN. En el caso de una elevada

degradación, escasez y fragmentación del ADN endógeno a la muestra, el ADN contaminante sería amplificado preferentemente, dado que éste es un ADN fresco, sin fragmentación ni daño molecular.

Las principales fuentes de contaminación proceden del personal que ha entrado en contacto directo con la muestra: arqueólogos, antropólogos, personal encargado de su conservación en un museo y personal del laboratorio. Por estas razones, son necesarias diversas medidas de prevención de la contaminación desde la exhumación del resto y durante el análisis en el laboratorio. Teniendo en cuenta que en los inicios de esta disciplina la errónea interpretación de los datos llevó a la publicación de resultados no reproducibles o contaminados, la comunidad científica comenzó a establecer unas medidas necesarias para la validación de resultados procedentes de muestras antiguas. Dichas medidas se han ido concretando poco a poco (Pääbo, 1989; Handt et al., 1994; Lindahl, 1997; Cooper y Poinar, 2000) y hoy en día representan unas pautas imprescindibles para la validación y publicación de resultados genéticos antiguos. Estas pautas, denominadas *criterios de autenticidad* (Poinar, 2003; Pääbo et al., 2004), se listan a continuación e incluyen las precauciones a tomar en cuanto a infraestructura, metodología, interpretación y reproducibilidad de resultados y estimaciones del estado de preservación biomolecular de la muestra.

1. Laboratorio exclusivo de ADN antiguo

Es muy importante que el análisis genético de restos antiguos se lleve a cabo en laboratorios especializados en la manipulación de muestras degradadas, puesto que la presencia de moléculas de ADN fresco en el ambiente podría llevar a una contaminación severa. Es necesario tener en cuenta que en el caso de que se trabaje con ADN antiguo de origen humano, existen múltiples fuentes de contaminación relacionadas con el entorno de análisis, no sólo por parte de los propios investigadores, sino de cualquier persona que haya podido entrar en contacto con las instalaciones y el material empleado en el laboratorio. Independientemente de la especie objeto de estudio, es imprescindible evitar el análisis de muestras recientes del mismo origen, puesto que la amplificación de moléculas perfectamente preservadas

aumentaría enormemente las probabilidades de contaminación de muestras antiguas. La contaminación ambiental incluye no sólo las moléculas flotantes en el laboratorio, sino también todo el material que se emplee, incluidos los aparatos y los reactivos.

2. Separación de las áreas de trabajo

Además del estudio de restos antiguos en laboratorios especializados y libres de contaminación adicional, es necesaria una separación entre las diferentes áreas de trabajo. Es muy importante separar las fases que preceden la amplificación del ADN de las que la siguen, puesto que en el segundo caso se trabaja con concentraciones de ADN mucho más elevadas, que podrían representar a su vez una fuente de contaminación. Por ello, estas dos fases se llevan a cabo en diferentes laboratorios, a ser posible localizados en edificios distintos o a cierta distancia. La fase más crítica es la primera, que va desde la recepción de la muestra hasta la extracción y amplificación de ADN. Es recomendable llevar a cabo estos primeros pasos en diferentes salas de un laboratorio, con control de contaminación y acceso restringido. En cambio, a partir de la amplificación del ADN es posible desarrollar los siguientes análisis en un laboratorio común, sin un control tan estricto de contaminaciones, puesto que ya se estaría trabajando con elevadas concentraciones de ADN no degradado.

3. Acceso restringido

Solo el personal especializado puede tener acceso a los laboratorios dedicados a la manipulación y análisis de muestras antiguas, sobre todo durante las fases de preparación de la muestra, extracción y amplificación. Además, es recomendable que la manipulación de la misma sea llevada a cabo durante todo el proceso por parte de un único investigador. Ambas medidas se ciñen al principio de reducción de posibles contaminaciones exógenas.

4. Material exclusivo

Todo el material y los reactivos necesarios para el estudio de muestras antiguas tienen que ser dedicados exclusivamente al análisis de muestras degradadas. En caso contrario, cualquier material o reactivo que haya entrado en contacto con ADN fresco, mejor preservado que el antiguo, representaría una fuente de contaminación severa para las muestras degradadas.

5. Empleo de controles o blancos

Durante las fases de extracción y amplificación de aDNA es muy importante el empleo de blancos o controles negativos. Se trata de mezclas de reacción que incluyen todos los reactivos necesarios para los experimentos en curso, y a los que no se añade ninguna fuente de ADN. Estos blancos se analizan en paralelo a las mezclas de reacción que contienen ADN y, para garantizar la autenticidad de los resultados, es necesario que no produzcan resultados positivos. En caso contrario, sería posible detectar contaminaciones puntuales o generalizadas en una determinada tanda de extracción o amplificación. Se puede tratar de contaminaciones procedentes de un reactivo, del investigador o de otras muestras analizadas en paralelo, así como de contaminaciones ambientales.

6. Descontaminación de la muestra

La manipulación de las muestras desde su llegada al laboratorio comienza con una limpieza superficial para conseguir la eliminación de posibles moléculas contaminantes en su exterior. Existen diferentes técnicas empleadas en esta fase que van desde un raspado manual hasta el empleo de lejía (Kemp y Smith, 2005). No obstante, existe otro procedimiento de naturaleza física que consiste en la eliminación de alrededor de medio milímetro de la capa superficial de la muestra mediante su abrasión con óxido de aluminio a presión dentro de una cámara dotada de presión positiva (Hummel, 2003). A continuación las muestras se someten a radiaciones ultravioletas, que crean dímeros de timina en el ADN localizado en su superficie. Estos compuestos impiden la

amplificación del ADN, de manera que los contaminantes remanentes o los que se hayan podido introducir durante la limpieza de la muestra quedan neutralizados.

7. Empleo de diferentes parejas de cebadores

El empleo de diferentes parejas de cebadores o *primers* para la amplificación de un mismo fragmento permite detectar y evitar la amplificación de pseudogenes insertados en el ADN nuclear (NUMTs) (Cooper y Poinar, 2000), así como posibles haplotipos quiméricos o contaminaciones.

8. Correlación inversa entre la longitud de los fragmentos amplificados y la eficiencia de amplificación

Resulta interesante monitorizar la longitud de los fragmentos amplificados en muestras antiguas, ya que normalmente no es posible recuperar secuencias con longitud superior a las 100-200 pb, debido a la fragmentación del ADN (Pääbo et al., 2004) y/o a la presencia de moléculas o lesiones inhibitoras (ver **Apartados I.3.2.2 e I.3.2.3**). Por ello, es posible observar una correlación inversa entre la longitud del fragmento amplificado y la cantidad de producto obtenido tras la PCR (Pääbo et al., 2004). La amplificación de fragmentos de diferente longitud en una misma reacción también permite rastrear posibles contaminaciones, puesto que solo se puede obtener amplificaciones de fragmentos largos en el caso de ADN bien preservado.

9. Comparación de perfiles genéticos con manipuladores y blancos

Para la correcta interpretación de los resultados y la exclusión de posibles contaminaciones por parte de los manipuladores es necesario comparar los perfiles genéticos obtenidos con los de aquellas personas que hayan podido entrar en contacto con las muestras o con las instalaciones.

La secuenciación de los blancos de extracción y de PCR que proporcionen resultados positivos y la comparación de las secuencias obtenidas en las diferentes tandas de

extracción y/o amplificación permiten identificar moléculas contaminantes procedentes de otras muestras analizadas en paralelo, de los reactivos o del ambiente de trabajo. Este tipo de contaminación se conoce como *carry-over*.

10. Sentido filogenético

Teniendo en cuenta que el ADN antiguo es un material degradado y fragmentado, es conveniente amplificar diferentes fragmentos solapantes, así como otras regiones informativas. El conjunto de resultados obtenidos tiene que presentar un sentido filogenético de acuerdo con la filogenia establecida para el correspondiente marcador en la población actual. Este criterio evita la consideración de secuencias quiméricas y permite la exclusión de los NUMTs.

11. Múltiples extracciones y amplificaciones

Este criterio está directamente relacionado con el principio de reproducibilidad de los experimentos científicos. Se deben realizar diferentes extracciones a partir de muestras diferentes de un mismo individuo y también múltiples amplificaciones a partir de un mismo extracto. Se trata de una medida indispensable para la identificación de contaminaciones introducidas durante el proceso experimental y de las *miscoding lesions* (ver **Apartado I.3.2.1**).

12. Análisis en un laboratorio independiente

Un ulterior criterio para la validación de los resultados consiste en la repetición de todos los análisis en un laboratorio independiente, desde la preparación de la muestra hasta su amplificación. Este criterio se basa también en el principio de reproducibilidad y se emplea para excluir contaminaciones masivas de un laboratorio específico. La condición ideal es que la muestra se envíe directamente desde la excavación o el museo a los dos laboratorios correspondientes.

13. Ensayos bioquímicos

Diversos trabajos realizados a inicios de los años noventa, pusieron de manifiesto que era posible inferir el grado de preservación del ADN a partir de la estimación realizada para otras moléculas, como el grado de racemización de ciertos aminoácidos (Poinar et al., 1996). Se consideraba que la tasa de racemización del ácido aspártico era equiparable a la tasa de depurinización de las moléculas de ADN. La ventaja de esta técnica sobre el análisis directo de los especímenes residía en que requiere una cantidad mínima de muestra, lo que justificaba no tener que destruir innecesariamente las muestras objeto de estudio si el valor de racemización obtenido no era compatible con la preservación del material genético. Sin embargo, recientemente se ha demostrado que no es posible extrapolar los datos de racemización para la estimación de la preservación del ADN, puesto que actúan diferentes mecanismos de degradación a nivel de los dos tipos de biomoléculas (Collins et al., 2009; Fernández et al., 2009).

14. Cuantificación

La cuantificación de las moléculas presentes en un extracto de ADN proporciona una medida de fiabilidad de los resultados obtenidos. Además permite estimar el número de amplificaciones o extracciones independientes a realizar: cuanto menos concentrado esté el ADN, más replicaciones serán necesarias para determinar su autenticidad. En el caso de que un extracto de ADN contenga un número de moléculas superior a 1.000, se considera que la identificación de la secuencia consenso podría realizarse a partir de dos amplificaciones independientes por extracto. En el caso de concentraciones inferiores, se recomienda replicar los resultados más veces (Pääbo et al., 2004).

15. Clonación bacteriana

La clonación bacteriana de los productos de amplificación se emplea para discernir entre las diferentes moléculas de ADN molde amplificadas en una PCR. Se basa en la inserción del ADN amplificado en un vector de clonación o plásmido, con el que se transforman células de la bacteria *E. coli*. El mecanismo de clonación hace que cada vector incorpore un único inserto y que cada célula bacteriana incorpore un único vector. Una vez dentro de la bacteria, ésta se multiplica reproduciendo el ADN incorporado como si fuera propio. De esta forma, cada colonia bacteriana, descendiente de la original, contendrá copias idénticas de cada molécula de ADN presente en la PCR. Alineando las secuencias obtenidas a través de la clonación de varios productos de amplificación procedentes de diferentes extracciones y amplificaciones de un mismo individuo, es posible identificar la secuencia endógena a la muestra, así como moléculas contaminantes y nucleótidos que han sufrido daño molecular o *Jumping PCR*.

I.3.3.4 INFLUENCIA DE LOS FACTORES AMBIENTALES Y DEL ENTORNO DE DEPOSICIÓN SOBRE LA PRESERVACIÓN DEL ADN ANTIGUO

Todas las reacciones de degradación y daño molecular descritas pueden ser más o menos rápidas en función de los factores ambientales a los que esté sujeta la muestra. El estado de preservación del ADN depende principalmente de las condiciones del entorno de deposición más que de su antigüedad (Nielsen-Marsh y Hedges, 2000; de Torres et al., 2002). El factor que se considera más determinante en la preservación del material genético es la temperatura y la historia térmica del ambiente de deposición (Smith et al., 2003). Los ambientes fríos con pocas fluctuaciones térmicas son los óptimos (Lindahl, 1993; Burger et al., 1999; Smith et al., 2001, 2003), puesto que frenan la degradación de las muestras (Höss et al., 1996). También la ausencia de agua y la baja humedad favorecen la preservación de ADN, dado que impiden el crecimiento de microorganismos y la degradación hidrolítica del material genético (Burger et al., 1999).

Asimismo, la radiación solar, y más especialmente su componente ultravioleta, y las radiaciones ionizantes perjudican la preservación del material genético no solo porque pueden provocar el aumento de la temperatura, sino también porque fomentan la formación de radicales libres que perjudican la integridad del ADN mediante procesos oxidativos (ver **Apartado I.3.2.2**) (Gilbert, 2006).

Por otro lado, los ambientes ácidos disuelven el fosfato de calcio degradando la hidroxiapatita de los restos óseos, mientras que condiciones alcalinas tamponan los efectos de los ácidos y favorecen la preservación de la estructura ósea. Los ambientes alcalinos favorecen la formación de bicarbonato, el cual aumenta la degradación de la hidroxiapatita (Bollongino et al., 2008). Como consecuencia, la preservación del material genético es mejor en entornos anóxicos, secos, oscuros, ligeramente alcalinos y caracterizados por bajas temperaturas.

I.3.3.5 LÍMITE DE PRESERVACIÓN DEL ADN ANTIGUO

Existe un límite máximo de preservación del ADN determinado experimentalmente (Pääbo y Wilson, 1991; Lindahl, 1993) y que ronda los 100.000 años en condiciones fisiológicas de pH y concentración de sales. Este límite actualmente está confirmado por la casi totalidad de los estudios de aDNA, siendo la única excepción que supera este límite un estudio de restos vegetales datados en 300.000 - 400.000 AP (Antes del Presente) (Willerslev et al., 2003).

En lo que concierne al género *homo*, el único espécimen que alcanza el límite de los 100.000 años de antigüedad es un hombre de Neandertal procedente de la cueva de Scladina (Meuse Basin, Bélgica) del que se pudo recuperar un corto fragmento de mtDNA (Orlando et al., 2006). Este límite puede ser más amplio en el caso de restos preservados en hielo o permafrost, condiciones óptimas para la preservación. En estos ambientes la preservación del material genético es excepcional, como demuestra la recuperación reciente del genoma casi completo de un individuo de 4.000 años a partir de un resto de cabello (Rasmussen et al., 2010). Asimismo, un ambiente desfavorable puede limitar enormemente la posibilidad de recuperación de ADN endógeno. En el

caso de papiros y restos humanos procedentes de zonas cálidas como Egipto, Marota y colaboradores (Marota et al., 2002) fijaron el límite máximo para la preservación del ADN por debajo de los 1.000 años. Sin embargo, recientemente se ha publicado un estudio ampliamente validado en el que se obtuvo información genética de STRs nucleares de momias egipcias con una antigüedad de unos 3.500 años, de la familia del faraón Tutankamon (Hawass et al., 2010).

A pesar de las mejoras tecnológicas introducidas en el campo del ADN antiguo durante los últimos 20 años, la probabilidad de recuperación de ADN endógeno sigue estando sujeta a las condiciones de preservación de los restos, siendo favorecida por climas fríos y secos frente a cálidos y húmedos.

I.4 APLICACIONES DEL ADN ANTIGUO

I.4.1 Estudios diacrónicos en Europa y la Península Ibérica, desde el Neolítico hasta la Edad del Hierro

El ADN antiguo representa una poderosa herramienta con la que sondear el pasado de las poblaciones y verificar hipótesis planteadas desde otras disciplinas como la arqueología, la antropología o la propia genética de poblaciones. Gracias al análisis directo de las poblaciones humanas antiguas es posible estudiar procesos a gran escala, como migraciones, expansiones poblacionales y cuellos de botella, así como continuidad o discontinuidad genética entre diferentes pobladores de una región, como en el caso de estudios diacrónicos. Para ello, es necesario integrar los resultados obtenidos con la información arqueológica, antropológica o genética disponible.

Entre las aplicaciones de esta disciplina destacan sus aportaciones a nivel del estudio del poblamiento de Europa, una de las cuestiones arqueológicas más debatidas durante las últimas décadas. A este propósito, se pretende estimar la contribución relativa de los diversos procesos migratorios acontecidos durante la prehistoria.

A continuación, se detalla el marco histórico de los períodos estudiados en la presente tesis doctoral mediante el análisis poblacional de restos antiguos que van desde el Neolítico hasta la Edad del Hierro. Asimismo, se presenta el conjunto de estudios genéticos actuales y antiguos presentes en la literatura acerca del poblamiento Europeo en general y peninsular en particular.

I.4.1.1 MARCO HISTÓRICO

Las primeras migraciones del hombre anatómicamente moderno (HAM) desde África hacia Europa, pasando por Próximo Oriente, se remontan al Paleolítico Superior, hace alrededor de unos 40.000 AP (ver **Figura I 7**) (Guy Straus, 1989). Hasta entonces

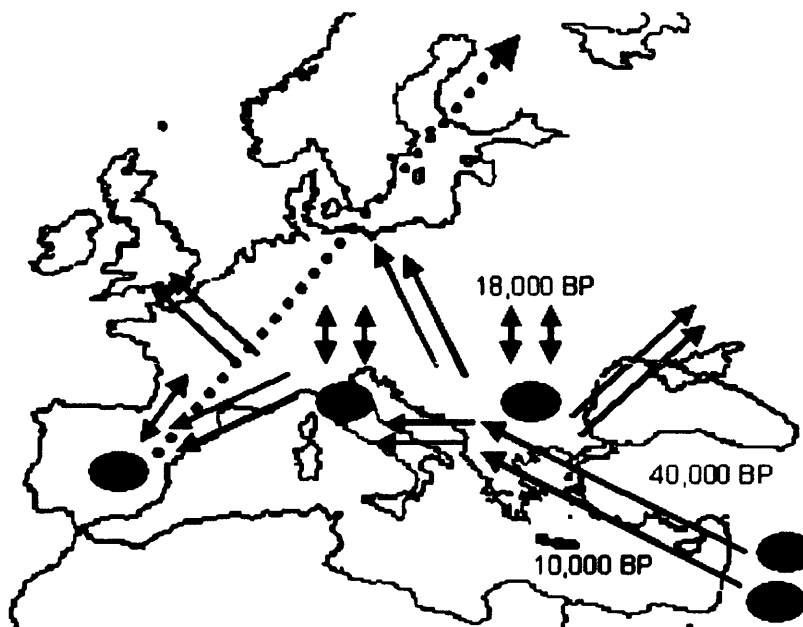
Europa estaba poblada solo por Neandertales, asociados a la industria lítica Musteriense (Eiroa, 2006). Los primeros restos de HAM se encontraron en África oriental, siendo fechados en aproximadamente 195.000 años (McDougall et al., 2005) y denominados inicialmente como “hombres de Cro-Magnon”, aludiendo a la localidad francesa en la que se describieron por primera vez (Quatrefages y Hamy, 1874). Los HAM presentaban una morfología craneal diferente y se han asociado a otro tipo de cultura material, el Auriñaciense (Condemí, 2011). Los HAM y los Neandertales convivieron durante unos 10.000 años, hasta la desaparición de los segundos hace unos 30.000 años. La relación entre ellos ha sido durante décadas el centro de un largo debate y hasta hace poco los datos genéticos, esencialmente basados en el estudio del mtDNA, no apoyaban la posibilidad de mezcla entre ambos (Krings et al., 1997; Ovchinnikov et al., 2000; Serre et al., 2004; Orlando et al., 2006; Briggs et al., 2009). Sin embargo, en 2010 se ha publicado el primer borrador del genoma de Neandertal (Green et al., 2010) y los resultados apuntan a que pudo haber cierta mezcla. Se ha estimado que la contribución genética del Neandertal al acervo genético moderno de descendencia no africana podría haber oscilado entre el 1 y el 4 %, apuntándose que probablemente se produjo mezcla en Próximo Oriente durante las primeras fases de expansión del hombre moderno hacia Europa (Green et al., 2010).

El siguiente evento importante a nivel demográfico ocurrió después del último máximo glacial (LGM, *Last Glacial Maximum*), hace unos 25.000 – 19.500 años, cuando la mayor parte de Europa estaba cubierta por los hielos y hubo una contracción de las poblaciones paleolíticas, que se concentraron en los llamados refugios paleolíticos (Clark et al., 2009). Éstos estaban localizados en las zonas sur-occidentales de Europa, esencialmente en la franja Franco-Cantábrica, en Italia, en los Balcanes y en las costas del Mar Negro (Dolukhanov, 1993; Semino et al., 2000; Gamble et al., 2004, 2005; Pala et al., 2009).

Solo a finales del Paleolítico, paralelamente a la retirada de los hielos, Europa se volvió a poblar a partir de dichos refugios (ver **Figura I 7**). Debido a la alteración climática, cambiaron las costumbres culturales y tecnológicas. Así las formas de arte se transformaron y se difundieron en Europa y en la Península Ibérica nuevos

instrumentos líticos, llamados microlitos por sus dimensiones (Eiroa, 2006). También cambió la dieta debido a la desaparición de grandes vertebrados. Este período se conoce como Epipaleolítico o Mesolítico y tuvo comienzo en Europa desde el 9.000 AP, durando hasta el comienzo del Neolítico (Eiroa, 2006).

Figura I 7 Esquema de los posibles movimientos poblacionales que tuvieron lugar en Europa durante la expansión del hombre moderno (en rojo), la repoblación de Europa a partir de los refugios paleolíticos (en morado) y la difusión de la agricultura (azul). Las flechas se refieren a los posibles movimientos poblacionales/culturales sugeridos por los datos arqueológicos, y los óvalos se corresponden con el comienzo de cada movimiento. Se señalan con flechas bidireccionales los movimientos de contracción demográfica en los refugios paleolíticos (óvalos morados), seguidos de una expansión poblacional después del último máximo glacial. La flecha discontinua representa la expansión post-glacial desde el refugio de la franja cantábrica hacia Europa septentrional (Barbujani y Goldstein, 2004).



Los cambios culturales y sociales más relevantes en la historia del hombre tuvieron lugar a partir del Neolítico, cuando la subsistencia humana pasó de una economía de recolección a una de producción, modificando sustancialmente el estilo de vida de las poblaciones paleolíticas (Price, 2000). El origen de la cultura Neolítica tuvo lugar en Próximo Oriente hace más de 10.000 años y se expandió hacia Europa en menos de 3.000 años (ver **Figura I 7**) (Price, 2000). Este cambio económico, social y cultural

estuvo caracterizado por la práctica de la agricultura y la ganadería, por el comienzo de la vida sedentaria en poblados y el empleo de materiales cerámicos y se suele denominar transición neolítica o neolitización.

Otros avances a nivel económico, cultural y social corresponden a las diferentes etapas de la llamada Edad de los Metales, cuando el empleo de instrumentos de piedra fue gradualmente sustituido por herramientas metálicas. Sus comienzos se remontan al V milenio AP y están asociados no solo al desarrollo de la metalurgia, sino también a importantes transformaciones culturales, tecnológicas y demográficas (Barandiarán et al., 2007). Esta etapa prehistórica se divide en tres períodos: la Edad del Cobre o Calcolítico, la Edad del Bronce y la Edad del Hierro. A continuación se describen las principales características de cada uno de los períodos prehistóricos que van desde el Neolítico hasta la Edad del Hierro.

I.4.1.1.1 EL NEOLÍTICO

La etimología de la palabra Neolítico se refiere a la Era de la piedra moderna, en contraposición a la Era de la piedra antigua, el Paleolítico. Sin embargo, ha ido adquiriendo muchísimos más significados, puesto que se ha empleado para definir un periodo cronológico, una fase cultural, un salto evolutivo, una forma de subsistencia, una estructura social y también una población (Price, 2000). Así pues, se han creado una serie de ambigüedades que han complicado más todavía la comprensión de esta fase (Dennell, 1992). Se suele hablar de “paquete neolítico” o *Neolithic package*, que incluye características específicas de la cultura y tecnología neolítica más que fechas concretas, dado que alcanzó las diferentes regiones en distintos momentos y además en muchos territorios las poblaciones neolíticas y mesolíticas convivieron durante largos periodos de tiempo. El paquete neolítico se refiere principalmente a los materiales líticos, las cerámicas, los poblados estables y los elementos tecnológicos típicamente asociados a las poblaciones agricultoras y ganaderas. Sin embargo, este paquete no llegó intacto en todas las áreas y en algunos casos se encontraron elementos de tecnología neolítica sin claras evidencias de una economía agro-pastoral.

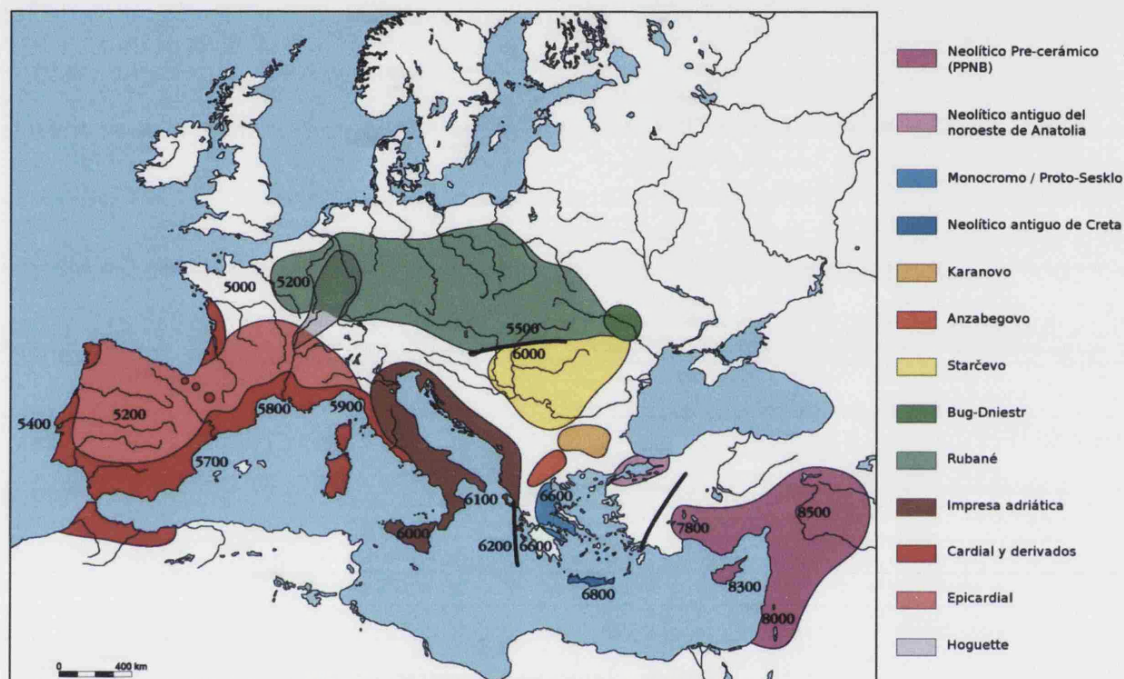
Además, este cambio tecnológico y cultural conllevó, aunque de manera gradual, un crecimiento demográfico importante, debido al aumento de recursos disponibles (Bocquet-Appel y Bar-Yosef, 2008). De hecho, al principio del Neolítico, se estima que la mortalidad aumentó, incluso respecto a los grupos de cazadores-recolectores. El cambio de dieta no resultó ser muy favorable, puesto que se basaba principalmente en el consumo de pocas especies y era muy dependiente de la cantidad de la cosecha. Se sospecha que la ganadería aumentó la probabilidad de desarrollo de enfermedades infecciosas transferidas de los animales a los hombres. Se encontraron pruebas de malnutrición y difusión de enfermedades a través del análisis antropológico de restos óseos de las primeras poblaciones neolíticas (Cohen, 1984). Sin embargo, una vez estabilizada, la población neolítica comenzó a crecer gracias a una serie de ventajas, entre las cuales estaban la posibilidad de almacenar alimentos en exceso y el requerimiento de intervalos menores entre nacimientos, gracias al sedentarismo.

La práctica agrícola surgió en diferentes lugares del planeta de forma independiente. Sus principales focos fueron Próximo Oriente, Mesoamérica y China. En nuestro continente, a finales del último máximo glacial (LGM, *Last Glacial Maximum*) los hielos se retiraron y el clima empezó a mejorar y a estabilizarse. Como consecuencia la población paleolítica que lo habitaba comenzó a crecer y a la vez empezaron a extinguirse los grandes vertebrados, como el mamut, probablemente por el cambio climático y el aumento de la presión de la caza. Aunque no se trate de fenómenos directamente relacionados, estos factores podrían haber empujado el desarrollo de la agricultura (Price, 2000). Las primeras comunidades agrícolas de Próximo Oriente no empleaban cerámicas y pertenecían a la cultura denominada PPN (*Pre-Pottery Neolithic*, Neolítico pre-cerámico). Durante esta primera fase acerámica la nueva economía se expandió desde Turquía hacia Macedonia y Grecia. A continuación se comenzó a emplear cerámica en todas las comunidades agrícolas. A partir de entonces se estimó que la cerámica podría representar un indicador arqueológico válido de la difusión de la agricultura (Price, 2000).

Su expansión hacia Europa fue relativamente rápida, ya que las primeras comunidades neolíticas se establecieron en el área Egea y en Grecia alrededor del 7.000 a.C. (a.C. antes de Cristo) mientras que en Inglaterra y Escandinavia no llegaron hasta poco

después del 4.000 a.C. (Price, 2000). De hecho, si se representan en un mapa las diferentes culturas neolíticas asociadas a su datación (Guilaine, 2003), es posible identificar un gradiente sureste noroeste, que refleja las fechas de expansión (ver **Figura 18**).

Figura 18 Mapa en la que se representan las dataciones de las diferentes culturas neolíticas, expresadas en años a.C. (imagen modificada a partir de Guilaine 2003).



En Europa se distinguen dos rutas principales asociadas a diferentes complejos cerámicos. Por un lado la ruta asociada a la cerámica lineal (*Linneerbandkeramic* o LBK), que se desarrolló en Europa Central, expandiéndose en varias etapas hacia el norte de Europa. Por otro lado, la ruta asociada a la cerámica Cardial, a lo largo de las costas mediterráneas, que incluyen Italia, sur de Francia y la Península Ibérica. En este último caso, las primeras comunidades neolíticas se desarrollaron principalmente a lo largo de las áreas costeras, expandiéndose después hacia el interior de la península (Martí Oliver, 2007).

Como ya se ha comentado anteriormente, la naturaleza de la expansión Neolítica ha sido y sigue siendo objeto de debate desde hace más de 50 años. Las primeras hipótesis arqueológicas implicaban una interpretación global del fenómeno neolítico y

las dos teorías contrapuestas se llamaron “difusión démica” y “difusión cultural”. La primera se refiere a un movimiento poblacional masivo con el completo reemplazamiento de las comunidades mesolíticas que habitaban Europa, mientras que la segunda sugiere que la cultura neolítica fue adoptada por los pobladores mesolíticos europeos, sin implicar movimientos poblacionales. En las últimas décadas se han desarrollado, sin embargo, hipótesis intermedias para explicar el fenómeno de difusión del Neolítico, que difieren en el grado de contribución genética y demográfica de las poblaciones neolíticas al sustrato mesolítico pre-existente. Zvelebil ha definido un total de siete modelos para la difusión del Neolítico con mayor o menor grado de mezcla (Zvelebil, 2001):

1. *Folk migration* (Migración de gentes): migración masiva con reemplazamiento.
2. *Demic diffusion* (Difusión démica): entendida como “ola de avance” (Ammerman y Cavalli-Sforza, 1984).
3. *Elite dominance* (Dominación de una élite): migración de una élite que impone un nuevo modelo cultural.
4. *Infiltration* (Infiltración): migración de un pequeño número de especialistas que se ocupan de una necesidad concreta de la población de acogida.
5. *Leapfrog colonisation* (Colonización intermitente): migración de pequeños grupos que se establecen en regiones óptimas, separados de la población indígena.
6. *Frontier mobility* (Movilidad fronteriza): intercambios entre agricultores y cazadores en las regiones de contacto.
7. *Regional contact* (Contacto regional): movimiento e intercambio de ideas.

La difusión del Neolítico en Europa podría explicarse bajo diferentes modelos en función de los mecanismos específicos que jugaron un papel importante en una región concreta. Los estudios de tipo local acerca del impacto neolítico se están consolidando

cada vez más, puesto que se trata de un fenómeno que engloba diferentes facetas culturales, como las citadas LBK y Cardial, que se producen en áreas determinadas y no en la totalidad del continente europeo.

En cuanto a la Península Ibérica, las primeras teorías acerca de la difusión del Neolítico promovían una influencia norteafricana mayoritaria (Bosch Gimpera, 1932), bajo la denominación de “Cultura de Cuevas”. Esta teoría estuvo en vigor durante mucho tiempo y fue apoyada por otros autores que la denominaron “hispano-mauritana” o “íbero-sahariana” (Bosch Gimpera, 1932; San Valero Aparisi, 1948). Fue unos años más tarde cuando se comenzaron a considerar las aportaciones orientales y del mediterráneo como imputables a un único movimiento cultural con diferentes facetas, todas asociadas a cerámica impresa (Bernabó Breá 1946). A partir de los años 60, se fue devaluando cada vez más la teoría de la influencia norteafricana en la península (Fletcher, 1963; Pellicer, 1964) y se comenzaron a considerar las conexiones culturales entre los dos continentes más bien como una emanación desde la esfera española hacia marruecos (Tarradell, 1965). En general, durante este período, las síntesis regionales se sustituyeron por las visiones globales de tipo difusionista (Manen et al., 2007).

Durante los años 80, se difundió otra corriente de pensamiento que proponía un modelo mixto para explicar el proceso de neolitización en la Península Ibérica, implicando una difusión démica a lo largo de las costas del mediterráneo así como la aculturación en las regiones más interiores. Se trata del “modelo dual”, por el cual los primeros grupos neolíticos costeros se difundieron progresivamente hacia el interior de la Península Ibérica englobando a los grupos epipaleolíticos, que a veces adquirieron solo algunas facetas del paquete neolítico (Fortea, 1973; Fortea y Martí, 1984; Bernabeu, 2002; Juan-Cabanilles y Martí, 2002).

Sin embargo, existen otras hipótesis más recientes, como la de Zilhão (2000, 2001) que interpretó los datos arqueológicos suponiendo una colonización marítima por parte de grupos pioneros, basándose en la rápida expansión del complejo Cardial en las costas ibéricas. A favor de esta hipótesis están la casi contemporaneidad y las homologías culturales entre grupos neolíticos de las costas mediterránea y atlántica y la neta

separación de los grupos mesolíticos a nivel territorial, cultural, tecnológico, alimenticio y funerario (Zilhão, 2001).

I.4.1.1.2 TRANSICIÓN A LA EDAD DE LOS METALES Y LA EDAD DEL COBRE

Así como ocurre para el Neolítico, la metalurgia del cobre aparece en diferentes momentos a nivel europeo y también peninsular (Rincón Martínez, 2007). Este primer período de la edad de los metales se desarrolló durante todo el III milenio a.C. y se denomina Edad del Cobre, Calcolítico o Eneolítico. Aunque represente un factor importante, el empleo de este metal representa sólo uno de los muchos rasgos culturales del Calcolítico (Delibes de Castro y Fernández Miranda, 1999). Durante el Calcolítico la metalurgia del cobre es incipiente con finalidades esencialmente de connotación decorativa o ritual, mientras que la industria lítica, principalmente basada en sílex, es abundante. Es solo durante la Edad del Bronce cuando disminuyen considerablemente los útiles líticos en favor de los metalúrgicos, cuya presencia y diversidad tipológica aumentan notablemente (Rincón Martínez, 2007).

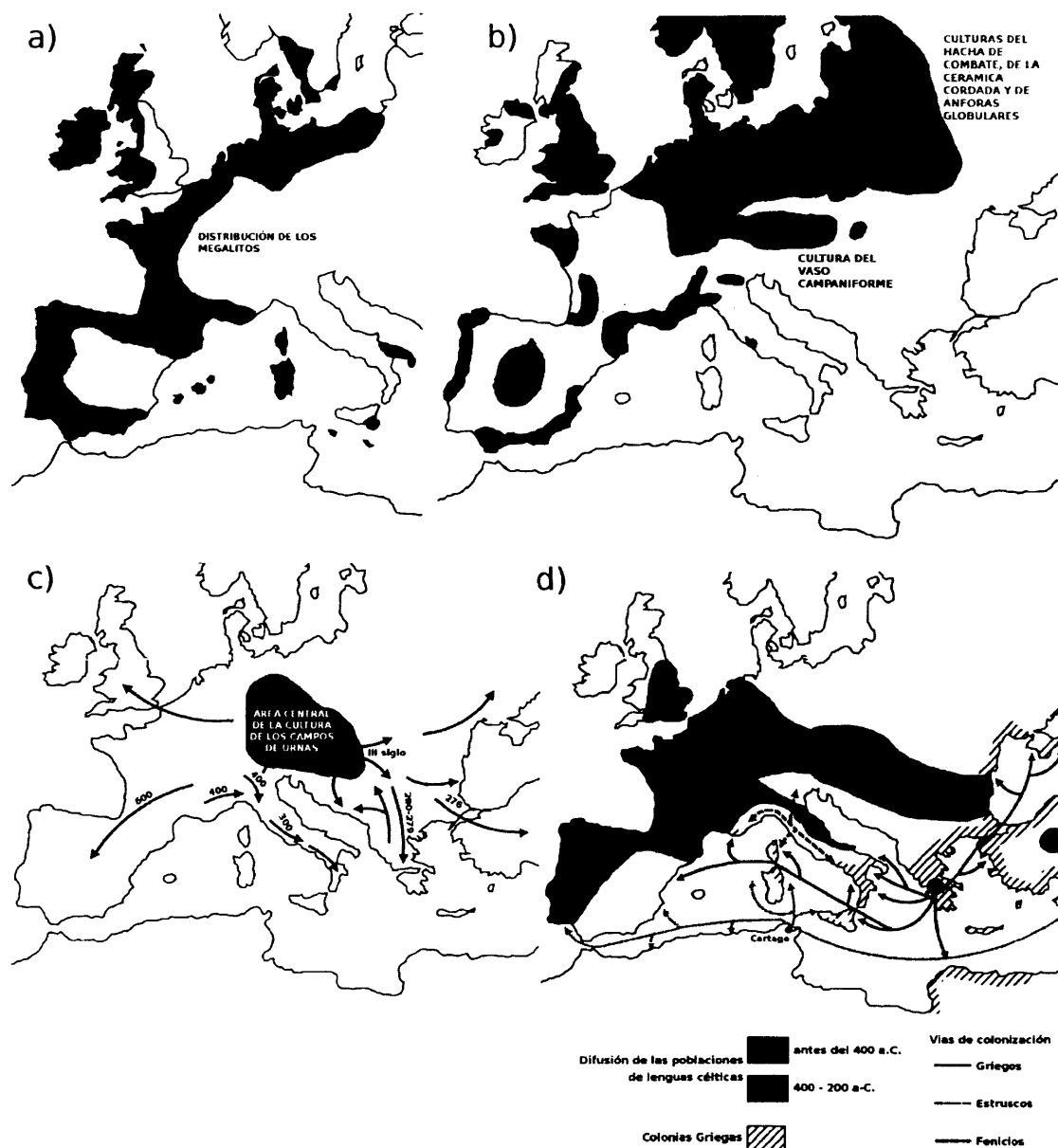
Desde finales del Neolítico y principios del Calcolítico se registra cierto cambio de hábitat, con el relativo abandono de cuevas en favor de poblados estables de mayor tamaño. Una de las características principales de esta época es el comienzo de la complejidad social, sugerida por diferentes aspectos arqueológicos. Por ejemplo, se han detectado ajuares desiguales dentro de grupos funerarios. Por otra parte, las diferencias de tamaño entre poblados y la fortificación de alguno de ellos sugieren la existencia de una jerarquía también a este nivel. Además existen evidencias arqueológicas del incremento en la producción agrícola y ganadera y del almacenamiento de excedentes. También hay evidencias de comercio de ámbar procedente de las regiones Bálticas y de marfil y ostras desde el norte de África (Jordá, 1998). Durante el Calcolítico se empiezan a desarrollar estructuras sociales cada vez más complejas que van desde jefaturas hasta verdaderos estados en la Edad del Bronce (Rincón Martínez, 2007).

Las primeras evidencias de fundición del cobre se han hallado en Anatolia y Kurdistán, regiones ricas en este mineral. Pero es en los Balcanes, durante el IV milenio a.C., donde se desarrolló plenamente la metalurgia calcolítica de manera autóctona. Aunque algunos elementos se expandieron hacia el resto de Europa mediante las redes de intercambio existentes, se han descrito diferentes desarrollos autóctonos de esta metalurgia. Por ejemplo, en la Península Ibérica se ha asociado el comienzo *in situ* de la fabricación del cobre con las culturas de Los Millares (Almería, España) y Vila Nova de São Pedro (Azambuja, Portugal) (Delibes de Castro y Fernández Miranda, 1999).

A nivel europeo, este período está caracterizado por diferentes culturas. Entre las más antiguas está la que se caracteriza por construcciones megalíticas, monumentos religiosos y funerarios contruidos con grandes bloques de piedra (ver panel **a** de la **Figura I 9**). Los más antiguos proceden de Inglaterra, donde se construyeron desde el V hasta el I milenio a.C. (Renfrew, 1987). La cultura megalítica también se difundió en casi toda la Península Ibérica a partir del Neolítico medio, caracterizó todo el Calcolítico y en algunos casos también llegó hasta la Edad del Bronce, con una complejidad arquitectónica creciente (Rincón Martínez, 2007).

Otra cultura que se difundió durante el Calcolítico final y el principio de la Edad del Bronce es la del vaso campaniforme (2.900-1.800 a.C.) (Muller y van Willigen, 2001), que ocupó gran parte de Europa central y occidental, incluida la Península Ibérica (ver panel **b** de la **Figura I 9**). Durante el siglo pasado, se han sucedido diferentes interpretaciones acerca de este extenso fenómeno cultural. Hasta los años 70 se consideró como una cultura homogénea que se expandió por toda Europa de mano de una élite de comerciantes y guerreros (Childe, 1930) y se propusieron diferentes posibles orígenes, entre los cuales se encontraba la Península Ibérica (Bosch Gimpera, 1919, 1940; Sangmeister, 1963). Sin embargo, esta cultura se solapaba con diferentes facetas en las distintas zonas de Europa, como la cultura de ánforas globulares en Europa central, lindando con la cultura de Baden y la más extensa cultura de la cerámica cordada y del hacha de combate en Europa central y septentrional (ver panel **b** de la **Figura I 9**) (Neustupny, 1963; Guilaine, 1966; Treinen, 1970).

Figura 1 9 a) Distribución de los monumentos megalíticos en Europa; b) Localización de algunas de las culturas de la Edad del Cobre y del Bronce: cultura del vaso campaniforme, de ánforas globulares, de la cerámica cordada y del hacha de combate; c) Difusión de la cultura de las culturas herederas de los Campos de Urnas (*Urnfield culture*) durante el primer milenio AP; d) Mapa de Europa durante el primer milenio a.C.: las influencias célticas y orientales (griegas y fenicias) (imágenes modificadas a partir de Cavalli-Sforza et al. 1994)



Fue solo a mediados de los 70 cuando se formuló una hipótesis alternativa basada en las cronologías absolutas, que identificaba Holanda como epicentro esta cultura (Lanting y van del Waals, 1974), desacreditando la hipótesis ibérica. Desde entonces el debate sigue abierto y la comunidad científica se ha centrado en las facies regionales, debido a la dificultad en identificar una verdadera homogeneidad y origen común. Los

elementos integrantes de esta cultura serían fruto más bien de las redes de intercambio que de la difusión de una cultura concreta (Garrido-Pena, 2007). Incluso en la Península Ibérica se han identificado estilos diferentes en correspondencia con las distintas regiones, como el Palmela en Portugal, el tipo Almeriano en Los Millares y el Continental en la meseta (Jordá, 1998).

A finales del Calcolítico y comienzos de la Edad del Bronce se difundió otra cultura en la parte occidental de las estepas rusas: la cultura de Yamna, caracterizada por una población pastoral nómada (Mallory, 1997). Se trata de una evolución secundaria de la economía agrícola, especialmente apta para ese tipo de medio ambiente, que se originó en Ucrania o en las montañas septentrionales del Cáucaso. En la parte occidental de las estepas, alrededor del III milenio a.C., se domesticó el caballo y se inventó el carro para transporte. A partir de entonces esta cultura comenzó su difusión en las áreas circundantes (Zvelebil, 1980), que está corroborada por la presencia de inhumaciones de túmulo (kurganes). Según la hipótesis de Marjia Gimbutas (Gimbutas, 1970), esta cultura es la principal candidata a ser la cuna del idioma proto-indoeuropeo. Esta autora interpretó su expansión desde una perspectiva bélica, de la que serían testigo los asentamientos fortificados y las inhumaciones diferenciales de los líderes guerreros. Sin embargo, otros autores suavizan esta connotación militar y tienden a destacar las diferencias entre grupos kurganes (Lull Santiago et al., 1991). Existen también otras perspectivas, que critican la hipótesis de asociación de la cultura kurgan con la difusión de las lenguas indoeuropeas, asociándolas a un origen más temprano en Anatolia (Gray y Atkinson, 2003) o incluso a la difusión de la agricultura (Renfrew, 1990).

I.4.1.1.3 EDAD DEL BRONCE

La edad del Bronce está caracterizada por el aumento de las fortificaciones y de las diferencias sociales a nivel individual, lo que se refleja en las diferencias entre ajuares en los enterramientos, que de colectivos pasan a ser individuales (Rincón Martínez, 2007). Las primeras poblaciones que empezaron a hacer aleaciones de cobre y estaño

obteniendo bronce, fueron las orientales, de Anatolia y de las islas griegas (Cíclades y Creta) entre el 3.200 y el 2.000 a.C. (Whitehouse y Whitehouse, 1977). Aquí se originaron las civilizaciones minoica, en Creta, y después micénica, en Grecia. La adopción de las técnicas necesarias para su elaboración fue facilitada por la red comercial existente y llegó hacia el 2.000 a.C. a Europa central, aunque aquí la hipótesis más probable es que la elaboración de este metal se desarrollara localmente (Cavalli-Sforza et al., 1994).

En la Edad del Bronce se suelen diferenciar tres fases en función de su cronología y sus características diferenciales y se emplea como referencia el centro de Europa: el Bronce antiguo (2.300-1.600 a.C.) ligado a la cultura de Unetice, el Bronce medio (1.600-1.250 a.C.) correspondiente a la cultura de los Túmulos y el Bronce final (1.250-800/750 a.C.) paralelo a la cultura de los Campos de Urnas (Rincón Martínez, 2007). En la Península Ibérica el Bronce Antiguo comienza hacia el 2.000 a.C. en el sureste, dónde se desarrolla la cultura argárica, que sigue a la civilización calcolítica de Los Millares y desde aquí se difunde al resto de la Península.

En el área egea la civilización minoica alcanza su auge durante el Bronce Medio. En Europa Central este período está asociado a la cultura de los túmulos, caracterizada por diferentes tradiciones funerarias respecto a la precedente cultura de Unetice; mientras que en la Península Ibérica es esencialmente una continuación del período precedente, el Bronce Antiguo (Rincón Martínez, 2007). Durante el Bronce Final ocurren los cambios más significativos de este período. Se corresponde con el pleno desarrollo de la civilización micénica en Grecia. En cambio, en Europa Central se desarrolla la cultura proto-céltica de los Campos de Urnas (*Urnfield culture*, ver panel c de la **Figura I 9**), que se difunde en todo el continente hasta el noreste de la Península Ibérica. Se trata de una cultura acompañada de nuevos hábitos funerarios, ligados a la incineración de los cuerpos. De todos modos, a pesar de ser agrupadas todas las áreas bajo el denominador común de la incineración, éstas son muy diversas, como consecuencia del sustrato que caracteriza a cada una (Rincón Martínez, 2007). Por ello, se considera que se trata más bien de una adopción de un rito que de la expansión de un pueblo (Fullola i Pericot y Nadal Lorenzo, 2005).

I.4.1.1.4 EDAD DEL HIERRO

Las sociedades de la transición entre la Edad del Bronce final y la del Hierro presentan un dinamismo creciente, paralelo al aumento del contacto entre diferentes culturas (Maya, 2007). Como ocurre con el desarrollo de las primeras prácticas agrícolas, el empleo del hierro se originó en Próximo Oriente y Grecia (siglo XII a.C.), mientras que en otras regiones europeas la aparición de sociedades con un nivel tecnológico comparable fue muy posterior (Waldbaum, 1978; Rieder y Wartke, 2009). En sus comienzos siguió predominando el empleo del bronce sobre el hierro, hasta que la elaborada tecnología asociada a la metalurgia del nuevo metal se perfeccionó. Llegó a Europa central alrededor del siglo X a.C. desde el Cáucaso y por vía terrestre. Hacia el IX siglo a.C. llegó por vía marítima a Italia desde la Hélade y Chipre mientras que algo más tarde, a través de intermediarios fenicios (procedentes de la parte más occidental del Creciente Fértil) y griegos llegó al Mediterráneo Occidental (Península Ibérica y Sur de Francia) (Eiroa, 2006).

Continuando con el paralelismo del Neolítico, tampoco hay acuerdo entre los autores acerca de la posible contribución poblacional asociada a la introducción de este metal (Maya, 2007). En el caso concreto de la Península Ibérica, se desconoce cómo y cuándo este metal fue introducido, aunque se suele relacionar con contactos con el exterior (Maya, 2007). Los pueblos de oriente, esencialmente fenicios, fundaron diferentes colonias en las costas peninsulares (como Gadir, la moderna Cádiz en España y sucesivamente Balsa, la moderna Tavira en Portugal) y tuvieron una influencia importante durante muchos siglos, desde el XI hasta el VII a.C. (ver panel **d** de la **Figura I 9**) (Pantoja, 2003). Gracias a estos nuevos contactos, se introdujeron cambios notables en esta época, que incluían diferentes avances tecnológicos, como el empleo del torno alfarero o de la forja del hierro, la obtención de plata por copelación y la modificación de la estructura urbanística, además de la introducción de ciertas especies domésticas como la vid, el olivo, los asnos y las gallinas, y el acceso a la escritura (Vega Toscano et al., 2003).

Los fenicios también fundaron Cartago, en las costas de Túnez, que llegaría a sustituir el dominio fenicio en el mediterráneo a partir del VI siglo a.C. (Blázquez et al., 1999). Fue entonces cuando desapareció una civilización que floreció localmente en la Península Ibérica (en Andalucía y Algarve), probablemente gracias a los contactos orientales. Se trataba de los tartesios, de la civilización de Tartessos, cuya existencia e importancia está descrita en textos griegos y latinos de época posterior, pero de la cual se desconocen muchos aspectos (Bermejo Tirado, 2007). A partir de entonces en la Península Ibérica se empezó a desarrollar la cultura ibérica, que perduraría en muchos casos hasta la época romana (la fecha de la victoria romana contra los cartagineses es el 218 a.C., durante la segunda guerra púnica) (Pantoja, 2003).

Durante la Edad del Hierro se suelen diferenciar dos zonas en la Península Ibérica, asociadas a dos áreas lingüísticas diferentes: la más occidental, caracterizada por lenguas célticas antiguas, y la oriental, en la que se emplean diferentes dialectos del ibérico. Se hace referencia a estos dos grupos con las denominaciones de celtíberos e íberos. En el primer caso se trataría de sociedades fuertemente influenciadas por la cultura centroeuropea de los Campos de Urnas y de sus herederas directas: la cultura de Hallstatt y sucesivamente la cultura de La Tène. Sin embargo, en el segundo caso, las influencias fenicia y oriental habrían tenido un impacto mayor (Vega Toscano et al., 2003). En su conjunto se trata de un mosaico de etnias caracterizadas por una organización social piramidal y un sistema político cada vez más complejo dominado por la aristocracia. Así surgieron unas élites que dominaban las redes comerciales y contaban con el empleo de elementos ideológicos y bélicos, como testifican los santuarios y la elevada presencia de armas en los ajuares de las clases dominantes (Bermejo Tirado, 2007). Otro aspecto que caracteriza estas culturas es el comienzo de la escritura. Sin embargo, aún no ha sido posible descifrar satisfactoriamente estas lenguas y las principales fuentes de documentación proceden de textos latinos y griegos, siendo éstas muy limitadas y a veces contradictorias (Bermejo Tirado, 2007).

I.4.1.2 DATOS GENÉTICOS DE LA POBLACIÓN MODERNA

En las últimas décadas se han desarrollado estudios de diferente índole para la comprensión del poblamiento de Europa, siendo una de las disciplinas más recientes la genética de las poblaciones humanas. Estos trabajos se han centrado esencialmente en las contribuciones paleolíticas y neolíticas al acervo genético europeo. Sus comienzos se remontan a los años 70, cuando el arqueólogo Ammerman y el genetista Cavalli-Sforza comenzaron un trabajo multidisciplinar que incluía datos arqueológicos, geográficos y, por primera vez, genéticos (Ammerman y Cavalli-Sforza, 1971, 1984; Menozzi et al., 1978). Analizaron conjuntamente las frecuencias alélicas de diferentes marcadores “clásicos” (grupos sanguíneos, antígenos del HLA y diferentes enzimas) empleando un análisis de componentes principales (PCA, Principal Component Analysis).

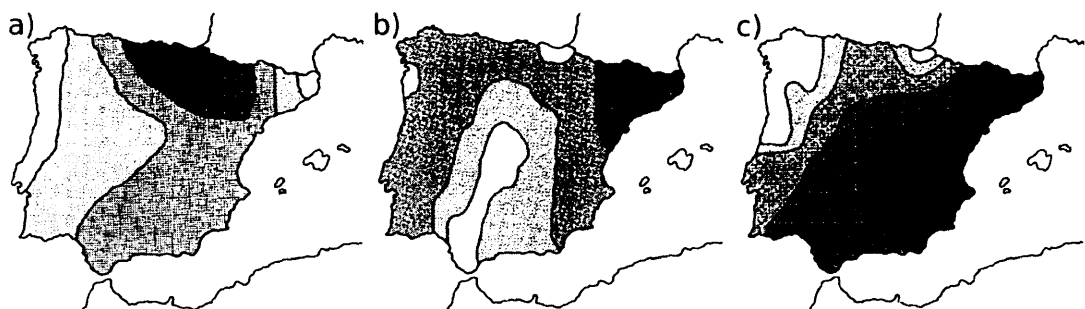
Los resultados fueron representados en un mapa geográfico de las regiones muestreadas, que incluía Europa y Próximo Oriente. La primera componente principal resumía el 28 % de la variabilidad total y en el mapa correspondiente era posible identificar una clina desde Próximo Oriente hacia el noroeste europeo. Estos resultados fueron interpretados como una prueba de una contribución démica importante durante la difusión de la agricultura en Europa. Estos autores aplicaron, además, el modelo matemático diseñado por Fisher (Fisher, 1937) conocido como “ola de avance” de la difusión neolítica, apoyando un modelo mixto de difusión con un elevado impacto démico pero con cierta difusión cultural. Hay que tener en cuenta, sin embargo, que los gradientes de frecuencias y las correspondientes clinas no están asociados a una datación concreta y que su asociación con un fenómeno demográfico específico, como la difusión neolítica, es hasta cierto punto arbitraria.

Otros movimientos poblacionales importantes siguieron la misma ruta, como los ocurridos durante la expansión del hombre moderno hacia Europa, la diáspora judía o la difusión de colonias griegas y fenicias durante la Edad del Hierro (ver **Apartados I.4.1.1 e I.4.1.1.4**). No obstante, mediante el test de Mantel, se encontró cierta correlación entre las distancias genéticas, geográficas y temporales correspondientes a la expansión neolítica (Sokal et al., 1991). Este test estadístico permite detectar

correspondencias entre matrices de distancias genéticas, geográficas y temporales. Al mantener constantes las distancias geográficas, las distancias genéticas presentaban correlaciones positivas con las dataciones del Neolítico en las diferentes regiones de Europa.

En lo que concierne a la Península Ibérica, los primeros estudios basados en marcadores clásicos también encontraron unas clinas específicas (Bertranpetit y Cavalli-Sforza, 1991). Pero en este caso la primera componente principal tenía un eje norte-sur (ver panel a de la **Figura I 10**) y solo la segunda seguía la clina europea este-oeste (ver panel b de la **Figura I 10**), asociándose a la expansión neolítica. Estos resultados se interpretaron como una mayor influencia paleolítica en Iberia, representada por la población vasca, que se localiza en el supuesto refugio paleolítico de la franja cantábrica a partir de la cual se expandió la población paleolítica después del LGM. También resulta interesante la tercera componente principal (ver panel c de la **Figura I 10**) que resalta la dicotomía entre las regiones atlánticas y mediterráneas, especialmente pronunciada durante la Edad del Hierro: por un lado las influencias célticas y por otro las orientales (ver **Apartado I.4.1.1.4**).

Figura I 10 Mapas sintéticos de la Península Ibérica en la que se representan las primeras tres componentes principales del estudio de Bertranpetit y Cavalli-Sforza (1991): a) primera componente principal, asociada a un 27 % de la variabilidad total, b) segunda componente principal, asociada a un 14 % de la variabilidad total, c) tercera componente principal, asociada a un 12 % de la variabilidad total. El intervalo entre el valor máximo y el mínimo de cada componente principal se ha subdividido en seis rangos iguales. La dirección de crecimiento de los valores de cada componente principal fue elegida de manera arbitraria (imagen modificada a partir de Bertranpetit y Cavalli-Sforza 1991).



Estudios más recientes basados en marcadores autosómicos (Chikhi et al., 1998; Barbujani y Bertorelle, 2001; Dupanloup et al., 2004; Belle et al., 2006) también identificaron una clina sureste-noroeste en Europa, que fue igualmente asociada a la

difusión del Neolítico. Asimismo se evidenció una huella neolítica importante en la Península Ibérica, aunque en menor medida respecto a otros países europeos, debido a la importante influencia paleolítica representada por la población vasca (Dupanloup et al., 2004).

El estudio de la distribución de los haplogrupos del cromosoma Y (Semino et al., 1996, 2000; Rosser et al., 2000) permitió cuantificar en alrededor del 20 % la contribución démica de los pobladores neolíticos, identificando los haplogrupos asociados al sustrato mesolítico y aquellos introducidos durante la expansión neolítica. El primer estudio poblacional del cromosoma Y (Semino et al., 1996) fue llevado a cabo sobre una muestra europea escasa. Aun así, fue posible identificar gradientes de frecuencias de dos polimorfismos: 12f2, que actualmente se corresponde con el haplogrupo J, y 49f, que se puede asimilar al actual haplogrupo R*[R1a]. El primero presentaba una frecuencia decreciente desde el sureste hacia el noroeste de Europa. El segundo, en cambio, presentaba una distribución de frecuencias inversa y estaba altamente representado en la población vasca. A pesar de la ausencia de análisis de clinas, los autores concluyeron que el primero de los gradientes apoyaba la hipótesis de difusión démica del Neolítico y el segundo se refería al sustrato paleolítico.

En el año 2000 se publicaron dos estudios (Rosser et al., 2000; Semino et al., 2000) que incluían un volumen de datos mucho más amplio. En el primer caso (Rosser et al., 2000) se identificaron dos haplogrupos (el J y el R*[xR1a, R1b8]) que presentaban clinas pronunciadas. En concreto, la frecuencia del haplogrupo J disminuía desde el sureste hacia el noroeste, mientras que la del R*[xR1a, R1b8] disminuía en sentido inverso (de noroeste a sureste). Estas clinas fueron interpretadas nuevamente como la contribución neolítica y paleolítica a la población moderna, apoyando una cierta difusión démica desde Próximo Oriente durante el Neolítico. En el segundo estudio (Semino et al., 2000), aunque no se testaron estadísticamente, se identificaron las mismas clinas, asociadas por un lado a haplogrupos “paleolíticos”, como el R y el I, y a haplogrupos “neolíticos”, el E3b, el F*[x I, J2, G, H, K], el G y el J2. Por ejemplo, en la Península Ibérica el J2 presenta picos de frecuencia en la zona noreste (Novelletto, 2007), donde llegaron los primeros pobladores neolíticos. Para reforzar estas hipótesis

de asociación, los haplogrupos citados fueron datados, encontrándose correspondencias entre dichas dataciones moleculares y las supuestas migraciones asociadas. Considerando el conjunto de haplogrupos “neolíticos”, los autores estimaron en un 22 % la contribución démica de la población neolítica. Sin embargo, no se tuvo en cuenta que la datación molecular de los haplogrupos simplemente indica de manera aproximada cuándo apareció esa variante en la población estudiada, y no tiene por qué estar altamente representada en la población de la época correspondiente.

Posteriormente se demostró que aumentando la resolución de los diferentes haplogrupos no era posible identificar las mismas clinas (Malaspina et al., 2001; Scozzari et al., 2001). El mismo conjunto de datos publicados por (Semino et al., 2000), fue posteriormente reanalizado por Chikhi y colaboradores (Chikhi et al., 2002) empleando un método basado en la simulación de genealogías, cuyo objetivo no consistía en identificar haplogrupos específicamente “neolíticos” o “paleolíticos”, sino en la estimación del impacto de las migraciones neolíticas definiendo solo las poblaciones parentales. Como representantes de las poblaciones neolítica y paleolítica se emplearon respectivamente datos genéticos relativos a las poblaciones de Próximo Oriente (Siria, Líbano y Turquía) y a la población vasca. Los autores calcularon las contribuciones relativas de estos dos grupos a la población europea actual, teniendo en cuenta también otros factores como la deriva genética. Los resultados revelaron un elevado impacto démico por parte de la población neolítica, con una contribución de al menos el 50 %, porcentaje mucho mayor que el inicialmente estimado por Semino y colaboradores (Semino et al., 2000).

Posterioros estudios poblacionales basados en el empleo del cromosoma Y demostraron las dificultades asociadas a este marcador, dado que ha sido caracterizado por importantes efectos fundadores a lo largo de la historia humana (Zerjal et al., 2003). Recientemente se han publicado otros estudios que tratan de interpretar el fenómeno de la neolitización en Europa mediante el empleo de otros marcadores del cromosoma Y. Entre ellos destaca el trabajo de Balaesque y colaboradores (Balaesque et al., 2010), en el que se estimó la frecuencia y la variabilidad haplotípica del haplogrupo R1b1b2-M269. Este haplogrupo presenta una

clina de frecuencia que aumenta de Próximo Oriente hacia Europa occidental con picos de frecuencia en el País Vasco. También presenta una paralela disminución de variabilidad a lo largo de este eje. Este gradiente fue tradicionalmente asociado con el sustrato paleolítico europeo, sin embargo en este trabajo se supuso un origen oriental y se interpretó como un marcador de la expansión neolítica. Esta conclusión fue apoyada por la pérdida de variabilidad de este haplogrupo debido a una expansión y a varios efectos fundadores a lo largo de su difusión. Otros estudios más recientes han criticado este trabajo, apuntando a un origen paleolítico y europeo del R1b1b2-M269 (Cruciani et al., 2011; Myres et al., 2011) o simplemente han resaltado las problemáticas de estas asociaciones, llamando a la cautela en la interpretación de los resultados (Busby et al., 2011). Más concretamente, Busby y colaboradores (2011) han destacado que, empleando un conjunto muestral lo suficientemente amplio, no es posible destacar gradientes de diversidad genética de este haplogrupo, como cabría esperar en el caso de que se tratara de un marcador de la expansión neolítica. Además, estos autores apuntan a la dificultad asociada a la datación de los haplogrupos, cuyas estimas podrían acercarse a valores más exactos empleando un conjunto de marcadores más amplios.

En conclusión, en lo que concierne al cromosoma Y, existen diferentes interpretaciones de los resultados. En general todos los estudios apuntan a la presencia de clinas, aunque muy controvertidas, que apoyarían una expansión démica desde Próximo Oriente en mayor o menor grado.

En lo que respecta al análisis del mtDNA, uno de los primeros trabajos en los que se abordó la cuestión de la contribución neolítica al acervo genético europeo fue el de Richards y colaboradores (1996). En este trabajo, se analizó la distribución y la frecuencia de haplotipos mitocondriales en un conjunto de poblaciones europeas y de Próximo Oriente. El análisis de los resultados sugería que la contribución de la población neolítica fue escasa, puesto que más del 80 % de los *clusters* (por aquel entonces aun no se empleaba el término haplogrupos) o linajes identificados presentaban una datación molecular que precedía a la expansión neolítica, habiéndose

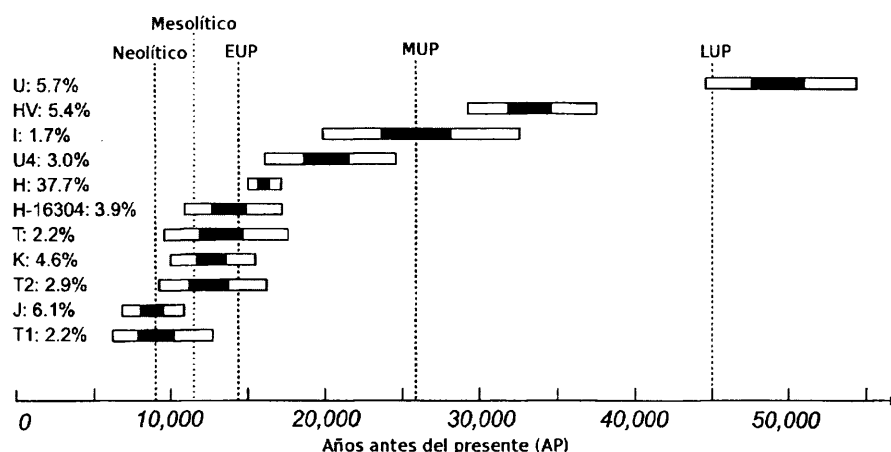
originado la mayoría de ellos durante el Paleolítico Superior final. Los linajes asociados a dataciones más recientes (correspondientes al Neolítico) estaban representados en tan solo el 12 % de la población europea y estaban esencialmente asociados al actual haplogrupo J. Estos resultados fueron interpretados otorgando una mayor importancia a la contribución de las expansiones post-glaciales durante el Paleolítico, sugiriendo una expansión neolítica esencialmente de carácter cultural. Además, la frecuencias de los linajes asociados al Neolítico estaban ligados a haplotipos ancestrales de Próximo Oriente pero no presentaban un gradiente sureste-noroeste.

A partir de la publicación de este trabajo (Richards et al., 1996), empezó un largo debate entre los genetistas de poblaciones. En primer lugar se criticó el mtDNA como marcador útil en el estudio de la historia humana, debido a que se trata de un marcador uniparental que solo refleja los linajes maternos y a los posibles efectos de la heteroplasmia sobre el cálculo de las tasas de mutación (Cavalli-Sforza y Minch, 1997). En este trabajo se volvieron a analizar los datos de Richards y colaboradores (1996) con las misma herramientas empleadas para el análisis de marcadores clásicos y los autores destacaron que ese conjunto de datos permitía simplemente detectar diferencias significativas entre el conjunto de poblaciones europeas y las de Próximo Oriente, sin permitir delinear conclusiones acerca del impacto del Neolítico (Cavalli-Sforza y Minch, 1997). La contestación en la que se defendía y justificaba la metodología empleada fue publicada inmediatamente (Richards et al., 1997). Surgieron otras críticas, como las relativas al hecho de que la edad molecular de un *cluster* no tiene por qué corresponder a la de la correspondiente población (Barbujani et al., 1998), que fueron sucesivamente contestadas por los autores aludiendo a que esta asociación solo se realizó después de la identificación de los haplotipos fundadores, a partir de los cuales se identificaron y dataron los correspondientes *clusters* (Richards y Sykes, 1998).

A continuación el equipo de Martin Richards publicó otros dos trabajos en los que se ampliaron las muestras y los marcadores analizados (Richards et al., 1998, 2000) y en los que se llegaron a conclusiones parecidas al primer estudio (Richards et al., 1996). En el último trabajo (Richards et al., 2000) se cuantificó una contribución neolítica de alrededor del 20 %, siempre basándose en la datación de los haplogrupos del mtDNA

mediante un análisis de haplotipos fundadores. En este artículo se estimó que del conjunto de haplogrupos mitocondriales europeos tan sólo los haplogrupos J, T1 y algunos haplotipos del U3, del X y del W podían asociarse a la expansión neolítica, puesto que su datación molecular era más reciente (**Figura I 11**). La conclusión principal de este trabajo apoyaba nuevamente una contribución fundamentalmente paleolítica al acervo genético europeo. De acuerdo con sus resultados, el haplogrupo más antiguo sería el haplogrupo U, que habría sido introducido en Europa al inicio del Paleolítico Superior (*Early Upper Palaeolithic* o EUP) por las expansiones del hombre anatómicamente moderno. Los haplotipos fundadores de los haplogrupos HV*, U1 y posiblemente también de U2 y U4 habrían llegado después, durante el Paleolítico Superior Medio (*Middle Upper Palaeolithic* o MUP). A finales del Paleolítico Superior (*Late Upper Palaeolithic* o LUP) habrían penetrado en Europa desde Próximo Oriente algunos sub-clusters del haplogrupo H.

Figura I 11 Datación molecular de los principales haplogrupos europeos (imagen modificada a partir de Richards et al. 2000).



Otros estudios han remarcado también la importancia de las migraciones Paleolíticas en detrimento de la difusión del Neolítico. Los macro-haplogrupos R0, U y JT se han relacionado con el poblamiento de Europa por parte del hombre moderno (Torroni et al., 2006), mientras que los haplogrupos V, H1, H3 y U5b se han identificado como los principales marcadores de la expansión post-glacial, debido a sus clinas de frecuencia,

más elevadas en los putativos refugios como la franja cantábrica (Torróni et al., 1998, 2001, 2006; Loogväli et al., 2004; Pereira et al., 2005; Alvarez-Iglesias et al., 2009).

Sin embargo, en un estudio reciente de García y colaboradores (García et al., 2011), se pone en discusión el rol protagonista de las expansiones post-glaciales, puesto que Europa central no se despobló completamente durante el LGM (Terberger y Street, 2002) y se afirma que antes de la expansión post-glacial seguían manteniéndose contactos entre Europa occidental y oriental (Svoboda, 2007). En este trabajo se apunta a una expansión post-glacial más regional y a una posible localización más amplia y meridional de los refugios paleolíticos.

En general, la mayoría de los trabajos publicados en los que se ha estudiado el poblamiento de Europa mediante el mtDNA apuntan a una difusión démica reducida durante la expansión neolítica. Sin embargo, se han publicado también diferentes trabajos con un enfoque más démico. En el primero (Simoni et al., 2000 a) se analizó una amplia muestra de poblaciones europeas y de Próximo Oriente, incluyendo a los Kurdos y los Drusos. El análisis de autocorrelación espacial apuntó a una baja estructuración geográfica asociada al mtDNA, sin embargo fue posible identificar ciertas clinas en la zona mediterránea, consistentes con la difusión del Neolítico. Sin embargo, en este trabajo no se discutieron las incongruencias existentes con otros marcadores, como el cromosoma Y. Hubo inmediatamente una crítica a este trabajo centrada en la definición errónea de ciertos haplogrupos, que habría influenciado el análisis de los resultados (Torróni et al., 2000). La crítica fue inmediatamente respondida aludiendo a que la definición de los haplogrupos se basaba en la secuencia de la región hipervariable y que el criterio de asignación empleado por otros autores no se pudo emplear como marco comparativo puesto que no se había hecho público (Simoni et al., 2000 b).

En un estudio posterior Richards y colaboradores (2002) aplicaron el método de análisis de PCA a una amplia muestra europea, caucásica y próximo oriental. Aquí fue posible identificar una clina sureste-noroeste, como nos esperaríamos bajo la hipótesis démica de difusión del Neolítico. Sin embargo, las poblaciones mediterráneas centrales y orientales, así como las de Europa del este, aparecían agrupadas con las de Próximo

Oriente, mientras que las de Europa central estaban agrupadas con las septentrionales. Teniendo en cuenta este resultado y que Europa central se considera una de las zonas de mayor migración neolítica, no se consideró como una evidencia a favor de la difusión démica (Richards et al., 2002).

En conclusión, a pesar del aumento de los trabajos basados en el análisis genético de la población moderna, existen resultados discordantes acerca de las contribuciones paleolíticas y neolíticas, en función del marcador analizado y de los diferentes métodos estadísticos empleados, traducándose en una comprensión limitada del poblamiento europeo (Soares et al., 2010). Por otra parte, aunque la transición neolítica probablemente represente un momento crucial en el modelado genético de los pobladores actuales de Europa, existen otras migraciones posteriores que habrían podido influenciar considerablemente la distribución de ciertas variantes genéticas. Sin embargo, existen pocos trabajos acerca de otros períodos, como los grandes movimientos poblacionales de la Edad del Hierro (ver **Apartado I.4.1.1.4**). Uno de los pocos trabajos centrados en migraciones post-neolíticas es el realizado por Zalloua y colaboradores (Zalloua et al., 2008), que se basa en el análisis del cromosoma Y. Este estudio permitió cuantificar en un 6 % la contribución fenicia a la población moderna en sus antiguas colonias mediterráneas, ocupadas durante la Edad del Hierro.

Otra cuestión poblacional muy interesante, al menos en lo que concierne a la Península Ibérica, es la contribución africana al *pool* genético actual. El papel del estrecho de Gibraltar como barrera para los movimientos poblacionales ha sido ampliamente discutido a la luz de los datos genéticos recuperados de ambos extremos del estrecho (Stringer y Andrews, 2005). El estudio del cromosoma Y apunta a cierta influencia africana en la Península Ibérica, probablemente ligada a la dominación musulmana en épocas históricas (Gérard et al., 2006; Capelli et al., 2009). Se ha estimado que la contribución norteafricana a nivel de linajes masculinos es de un 7.7 % (Capelli et al., 2009).

También el mtDNA es testigo, puesto que la frecuencia del haplogrupo U6, típicamente norteafricano, es mayor en Iberia que en el resto de Europa (Plaza et al., 2003; Pereira

et al., 2005). Sin embargo, es difícil discernir si este haplogrupo representa la expansión islámica medieval o bien tiene un origen más antiguo (González et al., 2003; Gonçalves et al., 2005).

También se han encontrado otras conexiones africanas con la península Ibérica, esencialmente asociadas al macro-haplogrupo mitocondrial L. Este linaje caracteriza las poblaciones subsaharianas y tiene una frecuencia muy reducida en Europa ($< 1\%$), pero alcanza frecuencias relativamente elevadas en la Península Ibérica (Pereira et al., 2005, 2010). No hay consenso acerca de la procedencia de este haplogrupo en Iberia. Ha sido asociado a episodios recientes, como el comercio de esclavos (Pereira et al., 2005) o incluso a migraciones prehistóricas (González et al., 2003).

También se han investigado posibles flujos poblacionales desde la Península Ibérica hacia el norte de África a nivel del mtDNA. Recientes trabajos apuntan a un impacto importante de las expansiones post-glaciales desde la Península Ibérica hacia las costas africanas, donde se encuentran haplogrupos cuyo origen ha sido tradicionalmente asociada al refugio paleolítico de la franja cantábrica (Rhouda et al., 2009). Entre ellos, el H1 y H3 son los sub-haplogrupos del H dominantes en la Península Ibérica y en el Norte de África, especialmente en la región noroeste, con frecuencias descendientes hacia Próximo Oriente (Ennafaa et al., 2009). El análisis del genoma mitocondrial completo de individuos ibéricos y norteafricanos pertenecientes a estos haplogrupos apuntan a un flujo génico prehistórico entre los dos continentes, debido a la elevada diferenciación genética entre los dos grupos poblacionales (Ennafaa et al., 2009; Ottoni et al., 2010).

Por otro lado, estos mismos haplogrupos (H1 y H3) también se han relacionado con eventos históricos, como la expulsión de los musulmanes, que habitaron la península ibérica durante diez siglos (entre el VIII y al XVII) (Cherni et al., 2009). La presencia de flujo génico entre los dos extremos del estrecho de Gibraltar, también se ha evidenciado a nivel de otras especies, como la que se considera como la primera especie de cabra domesticada (*Capra hircus*), cuya variabilidad mitocondrial y del cromosoma Y apunta a la presencia de mezclas a nivel de ambos lados del estrecho (Pereira et al., 2009).

Sin embargo, un estudio más reciente (Currat et al., 2010) que incluye la información genética de marcadores uniparentales (mtDNA y cromosoma Y) así como varios sistemas autosómicos (ABO, RH, GM, MNSs y HLA-DRB1), ha permitido identificar la presencia de flujo génico (sobre todo a nivel del cromosoma Y) a través del estrecho de Gibraltar desde épocas pre-neolíticas. En este trabajo también se evalúa el impacto de migraciones masculinas y femeninas diferenciales así como el efecto de la selección natural sobre uno de los marcadores nucleares estudiados (HLA-DRB1), que habrían podido distorsionar la interpretación de los resultados en precedentes estudios (Currat et al., 2010).

Para una mejor comprensión de la historia genética de nuestro continente es necesario estudiar *in situ* los posibles efectos de diferentes movimientos poblacionales a lo largo de nuestra historia. Una disciplina que permite comparar las hipótesis planteadas a partir de la genética actual es el estudio directo del ADN antiguo, cuyas contribuciones se describen a continuación.

I.4.1.3 DATOS GENÉTICOS DE LAS POBLACIONES ANTIGUAS

A pesar de las limitaciones asociadas a este tipo de estudios, como la degradación del material genético, la susceptibilidad a la contaminación y el número reducido de muestras, el ADN antiguo puede proporcionar datos de primera mano para el estudio de los orígenes de las poblaciones humanas.

En el caso de la dispersión del Neolítico en Europa, existen diferentes maneras de enfocar su análisis a partir de datos genéticos antiguos. En primer lugar, es posible estudiar la composición genética de los primeros pobladores neolíticos europeos y compararla con la población actual de las mismas regiones, para destacar afinidades y divergencias. Es interesante por otra parte, el estudio diacrónico de los pobladores de los períodos anteriores y posteriores al Neolítico en una misma región. De esta manera es posible evaluar la continuidad o discontinuidad genética entre el Mesolítico y el Neolítico. Además, resulta muy interesante la caracterización de las primeras

poblaciones neolíticas de Próximo Oriente, en conexión arqueológica con las primeras expansiones neolíticas.

La región en la que se dispone de una mayor cantidad de datos genéticos de muestras neolíticas antiguas es Europa central. A esta región llegó hace 7.500 años una de las principales culturas del Neolítico antiguo, la cultura LBK (ver **Apartado I.4.1.1.1**). Hasta la fecha se han caracterizado 41 muestras neolíticas alemanas y una austríaca pertenecientes a dicha cultura (Haak et al., 2005, 2010) y 20 muestras de grupos de cazadores-recolectores de las regiones circundantes (Alemania, Lituania, Polonia, Rusia, Holanda) (Bramanti et al., 2009). A pesar del escaso solapamiento geográfico de los dos períodos, entre las principales conclusiones de estos trabajos está la identificación de cierta ruptura genética entre los dos períodos (Bramanti et al., 2009), que parece apuntar a que junto con la cultura neolítica se produjo un movimiento poblacional importante.

En lo que concierne a la muestra neolítica, se han estimado ciertas afinidades con la población actual de Próximo Oriente, apuntando a una contribución démica importante durante la expansión neolítica desde esta región (Haak et al., 2010). Sin embargo, la población neolítica centroeuropea presenta una frecuencia muy elevada del haplogrupo N1a (~17 %), que está muy poco representado en la población actual europea (0,2 %). Inicialmente este haplogrupo fue propuesto como un marcador específico de la expansión neolítica (Haak et al., 2005), ya que tampoco fue identificado en muestras mesolíticas (Bramanti et al., 2009). Esta hipótesis ha sido apoyada por la presencia del mismo en otra muestra neolítica procedente de un enterramiento megalítico francés (Deguilloux et al., 2011 b). Recientemente se ha reconstruido la filogeografía del haplogrupo N1a a partir de secuencias mitocondriales completas (Palanichamy et al., 2010) y los resultados apuntan a que este linaje se ha originado en diferentes poblaciones, que incluyen Europa del este y central, así como Próximo Oriente, sugiriendo una expansión neolítica de pequeños grupos en un proceso conocido como “leapfrog colonisation” (Zvelebil, 2000) (ver **Apartado I.4.1.1.1**), más que una verdadera expansión démica. Rowley-Conwy apoya esta interpretación de los resultados proponiendo un origen paleolítico del haplogrupo N1a entre grupos balcánicos, griegos o del Mar Negro (Rowley-Conwy, 2009).

Otra región en la que se ha estudiado la transición Mesolítico-Neolítico es el norte de Europa (Malmström et al., 2009). También en este caso se detectó discontinuidad genética entre los dos períodos estudiados, sin embargo, se trataba de dos poblaciones de cronología mucho más tardía al LBK y prácticamente contemporáneos, asociados cada uno de ellos a una economía mesolítica o neolítica. Además, la disparidad del tamaño muestral entre los dos períodos estudiados (19 de la cultura *Pitted Ware*, de tradición mesolítica, datados en 4.800 - 4.000 AP, y solo 3 de la cultura de los vasos de embudo, con elementos neolíticos, datados en 5.500 - 4.500 AP) probablemente haya segado los resultados. Esta región presenta un contexto diferente respecto al resto de Europa, puesto que aquí la transición neolítica fue más tardía, resultando el modo de vida mesolítico más apto para la supervivencia en estas regiones.

Es importante destacar que en las dos muestras pre-neolíticas de Europa central y septentrional el haplogrupo U5 se encuentra altamente representado (Bramanti et al., 2009; Malmström et al., 2009). Estos datos, junto con el análisis filogenético a partir de secuencias completas de mtDNA, sugieren que este haplogrupo se habría difundido en el territorio europeo como consecuencia de la recolonización del norte de Europa a partir de los refugios paleolíticos del sur, tras el LGM (Malyarchuk et al., 2010).

Recientemente, se han aportado datos acerca de la composición genética neolítica de una sur de Francia (Lacan et al., 2011). Se trata de un conjunto de 29 muestras procedentes de una necrópolis del Neolítico Final, de las que se obtuvieron perfiles de mtDNA y de cromosoma Y. Los resultados apuntan a un diferente origen de los linajes masculinos y femeninos, puesto que el mtDNA presentaba un 80 % de variantes tradicionalmente asociadas al sustrato paleolítico, mientras que el cromosoma Y presentaba una composición genética similar a los actuales habitantes mediterráneos del sur de Europa (como Turquía, Chipre, Portugal e Italia). La ausencia de variantes de Cromosoma Y y de mtDNA relacionadas con la expansión del Neolítico desde Próximo Oriente en la muestra analizada (haplogrupos R1b-S116 y N1a respectivamente) fue interpretada como evidencia de la difusión del Neolítico a esta región a través de la ruta mediterránea. Sin embargo, no hay que olvidar que para interpretaciones a gran

escala serían necesarios más datos y en este trabajo se ha estudiado una sola necrópolis. Además, estas muestras tienen una datación muy posterior respecto a los primeros pobladores neolíticos, por lo tanto el *pool* genético y las tradiciones culturales pudieron cambiar a lo largo de esos milenios.

En el caso de la Península Ibérica los datos de ADN antiguo de época prehistórica disponibles, hasta el desarrollo de este trabajo de tesis, son escasos. Entre los primeros estudios, está el análisis por RFLPs (*Restriction Fragment Length Polymorphisms*) de un amplio conjunto de muestras prehistóricas del País Vasco (Izagirre y de la Rúa, 1999) en el que se determinaron haplogrupos mitocondriales en 121 muestras comprendidas entre el Neolítico y la Edad del Bronce. Una de las observaciones más interesantes de este estudio es la ausencia del haplogrupo V en muestras prehistóricas vascas. Este haplogrupo se encuentra altamente representado en la población actual vasca y, a partir de estudios de genética actual, fue propuesto como uno de los marcadores de la expansión post-glacial desde los refugios paleolíticos de la franja cantábrica (Torroni et al., 1998, 2001, 2006; Loogväli et al., 2004; Pereira et al., 2005; Alvarez-Iglesias et al., 2009) (ver **Apartado I.4.1.2**). Sin embargo, su ausencia en períodos prehistóricos posteriores (Neolítico – Edad del Bronce) apunta a otros fenómenos, como la deriva genética, como responsables de la frecuencia y distribución actual del haplogrupo V. El trabajo de Izagirre y de la Rúa (1999) señaló por primera vez que los datos genéticos de poblaciones actuales deben ser interpretados de forma cautelosa, poniendo además de manifiesto la importancia de la comprobación *in situ* de hipótesis arqueogenéticas.

Otras muestras prehistóricas de la Península Ibérica estudiadas mediante técnicas de ADN antiguo proceden de dos poblaciones de Cataluña. Se trata de un conjunto de 11 especímenes del Neolítico Medio (5.500-5.300 AP) (Sampietro et al., 2007) y 17 muestras íberas catalanas (Sampietro et al., 2005). A diferencia de los resultados obtenidos en Europa central, los datos de las poblaciones catalanas apuntan a la presencia de continuidad genética para ambos períodos con la población moderna (Sampietro et al., 2007). Estas conclusiones difieren de las observaciones

arqueológicas en estas regiones, que sugieren un modelo mixto, conocido como “modelo dual” (Fortea, 1973; Fortea y Martí, 1984; Bernabeu, 2002; Juan-Cabanilles y Martí, 2002) o un modelo de colonización pionera (Zilhão, 2000, 2001), con aportes demográficos respectivamente más o menos pronunciados de los primeros grupos neolíticos que llegaron a las costas catalanas (ver **Apartado I.4.1.1.1**).

Hay que tener en cuenta, sin embargo, que las muestras neolíticas catalanas analizadas no corresponden a los primeros pobladores neolíticos de la Península (Cultura del Neolítico Cardial, ver **Apartado I.4.1.1.1**), por lo que los resultados podrían estar sesgados por causa de los procesos demográficos ocurridos entre la época Cardial y el Neolítico Medio.

La continuidad genética con la población moderna se ve reflejada también en el estudio de los pobladores íberos catalanes (Sampietro et al., 2005), que presentan una composición genética muy parecida a la actual. Para la interpretación de estos resultados, hay que tener en cuenta las características socio-culturales de esta población, cuyos rituales funerarios estaban asociados a la cremación (ver **Apartado I.4.1.1.4**). Las muestras analizadas pertenecían a aquella pequeña proporción de población que tuvo un tratamiento funerario diferencial, sin incineración. Por ello, los resultados obtenidos a partir del análisis de esta minoría podrían estar sesgados.

En su conjunto, los resultados de ADN antiguo de épocas prehistóricas disponibles hasta el momento son bastante parciales, debido a las limitaciones de las muestras disponibles y de las técnicas. Sin embargo, la información proporcionada está contribuyendo de manera significativa a la comprensión de fenómenos históricos y prehistóricos.

I.4.2 Relaciones familiares e Identificación

El empleo de datos genéticos en el estudio de relaciones familiares y en la determinación molecular del sexo, puede aportar información única y muy valiosa a diferentes áreas de conocimiento como la arqueología, la antropología o las ciencias forenses. Entre sus aplicaciones se encuentra, por ejemplo, la identificación de personajes históricos mediante la comparación con descendientes vivos o muertos, contribuyendo a la comprensión de la historia y de la prehistoria. En este sentido, pueden servir para corroborar las hipótesis formuladas desde la genealogía o la antropología.

Por otra parte, la comprobación de relaciones familiares cercanas en enterramientos múltiples resulta una aplicación muy común del ADN antiguo en el campo de la arqueología. Los resultados de estos estudios pueden tener consecuencias importantes en la comprensión de los hábitos sociales de las culturas antiguas. El análisis de las variantes genéticas en un contexto arqueológico puede permitir comprender importantes transiciones sociales, como la que ocurrió durante el Calcolítico, cuando aumentó la estratificación social. Mediante la genética se podría comprender si la base de la jerarquización estaba relacionada con una élite de personas biológicamente relacionadas.

A la hora de llevar a cabo un estudio de relaciones familiares en muestras antiguas hay que emplear diferentes marcadores, en función de la edad de la muestra y del tipo de relación que se desee estudiar. Por una parte, el estudio de marcadores de linaje como el mtDNA o el Cromosoma Y permite explorar relaciones familiares a lo largo de un linaje. La ventaja del uso del mtDNA en el establecimiento de relaciones familiares en muestras antiguas reside en su mayor abundancia, lo que hace más probable su recuperación en muestras antiguas o muy degradadas. En este caso, el hecho de que dos individuos compartan el mismo haplotipo mitocondrial no significa necesariamente que se trate de familiares cercanos, sino que a lo largo de su genealogía compartieron un ancestro mitocondrial común. En casos de tumbas múltiples, resulta de gran interés para la arqueología conocer la relación familiar entre

ellos. En este contexto, el mtDNA permitiría excluir una hipótesis de maternidad si se presentaran dos haplotipos diferentes, pero en la situación opuesta, sería difícil demostrar dicha relación debido a la reducida variabilidad de este marcador. Pues sería necesario el estudio en paralelo de otros marcadores para aumentar la probabilidad de maternidad estimada.

A su vez, es posible emplear el Cromosoma Y para la identificación de relaciones familiares a lo largo de la línea paterna. La principal limitación de este marcador es su escasez, lo que dificulta su recuperación en muestras muy degradadas. Sin embargo, la variabilidad de algunos marcadores situados en el cromosoma Y, como los STRs, es elevada y, a diferencia del mtDNA, ofrece un elevado poder discriminante en las relaciones de parentesco.

Contrariamente a lo que sucede con otros marcadores nucleares, no existen *kits* comerciales especialmente diseñados para amplificar marcadores del cromosoma Y en muestras muy degradadas. Sin embargo, en los últimos diez años, se han publicado diferentes trabajos en los que ha sido posible obtener información de marcadores del cromosoma Y a partir de individuos bien preservados, mediante diferentes diseños experimentales. Entre los especímenes analizados de mayor antigüedad destacan aquellos preservados en permafrost (Bouakaze et al., 2007; Kemp et al., 2007; Keyser et al., 2009; Rasmussen et al., 2010) y algunos individuos neolíticos centroeuropeos (Haak et al., 2008, 2010) y franceses (Lacan et al., 2011). Por otro lado, muestras de origen más reciente han sido caracterizadas con éxito, como es el caso de individuos de época medieval (Csányi et al., 2008; Vanek et al., 2009).

Los marcadores de linaje, sin embargo, no permiten precisar el grado exacto de relación genética entre individuos. Para ello es necesario estudiar diferentes marcadores localizados en los cromosomas autosómicos, de los cuales se hereda una copia por progenitor. Para estos marcadores se han desarrollado *kits* de amplificación especialmente diseñados para ADN degradado, que incluyen una mezcla de reacción con elevada eficiencia y una batería de cebadores que amplifican fragmentos cortos.

Otro marcador muy empleado en el campo del aDNA es el gen de la amelogenina, localizado en los cromosomas sexuales X e Y, y empleado comúnmente para la

determinación molecular del sexo. Este gen ha sufrido una delección de 6 pb en el cromosoma X a lo largo de la evolución. La amplificación por PCR de la región que contiene la delección permite diagnosticar mediante técnicas electroforéticas el sexo del individuo. Se trata de una metodología estándar, que actualmente está incluida en todos los *kits* comerciales de PCR empleados para la identificación. Gracias al desarrollo de estas técnicas, hoy en día es posible determinar el sexo de individuos subadultos así como de esqueletos muy incompletos, cosa que no es posible empleando exclusivamente metodologías morfométricas.

Otro marcador potencialmente útil para el estudio de relaciones familiares debido a su patrón de herencia es el Cromosoma X. En este caso, uno de los dos cromosomas X de la madre se transmite a los hijos varones, mientras que las hijas lo reciben del padre. La información contenida en el cromosoma X puede resultar de gran interés en el estudio de relaciones de parentesco en los que no esté disponible la información de ambos progenitores. Las aplicaciones del Cromosoma X en el campo del aDNA, por otra parte, resultan muy limitadas, puesto que su análisis rutinario en el campo forense está aún en sus comienzos y actualmente no están disponibles *kits* comerciales específicamente diseñados para muestras muy degradadas.

La elección del marcador a estudiar está condicionada por el tipo de estudio que se desee llevar a cabo. Como ya se ha mencionado, para evaluar relaciones cercanas, como padre e hijo, se obtendrán probabilidades mayores empleando marcadores nucleares. Si en cambio se trata de relaciones lejanas, como las que existen en los estudios de identificación de personajes históricos, estos marcadores pierden utilidad puesto que se recibe la mitad de la información genética de cada uno de los progenitores y en cada generación ocurre recombinación entre ellos. Por ejemplo, una persona hereda la mitad de la información genética de cada uno de sus padres, pero un cuarto de cada abuelo, un octavo de los bisabuelos, etc. Así pues en estos casos es necesario estudiar marcadores de linajes paterno o materno siempre que existan descendientes vivos o muertos más recientes.

El estudio de relaciones familiares mediante la genética molecular se ha aplicado principalmente a restos históricos. Sin embargo, existen algunos trabajos en los que fue posible analizar restos más antiguos. El primer trabajo que empleó diferentes marcadores genéticos para el análisis de relaciones familiares a partir de un conjunto de 62 muestras protohistóricas es el que llevaron a cabo Keyser-Traqui y colaboradores (Keyser-Tracqui et al., 2003), en el que se consiguieron analizar el 74 % de las muestras para mtDNA el 79 % para nuDNA y el 77 % de los individuos masculinos para más de tres marcadores del cromosoma Y. Los resultados obtenidos permitieron comprender los hábitos funerarios, así como las relaciones sociales entre diferentes sectores de la necrópolis, con implicaciones arqueológicas relevantes.

Algo más antiguas son las muestras estudiadas por Haak y colaboradores (Haak et al., 2008). Se trata de enterramientos múltiples de individuos pertenecientes a la cultura de la cerámica cordada (4.600 AP, Eulau, Alemania). Gracias al análisis genético de diferentes marcadores (mtDNA, nuDNA y cromosoma Y) fue posible identificar relaciones familiares entre individuos adultos e infantiles enterrados conjuntamente. En base a la información genética obtenida, analizada paralelamente a los datos arqueológicos, las conclusiones de este estudio apuntan a que se trataba de una sociedad patrilocal y exogámica, en la que las relaciones familiares jugaban un papel muy importante.

Otro estudio en el que se analizaron las características de una sociedad antigua, a partir de restos humanos del Neolítico Final es el de Lacan (Lacan et al., 2011). En este caso, se analizaron marcadores del mtDNA y del cromosoma Y, que permitieron inferir ciertas relaciones genéticas a lo largo de la línea paterna entre individuos enterrados conjuntamente, apuntando a que se tratara de una sociedad patrilocal, donde los hombres se quedaban en el lugar de nacimiento mientras que las mujeres podían proceder de las regiones circundantes.

Entre los más antiguos merece la pena destacar que Bramanti y colaboradores (2009) obtuvieron nuDNA en tres de las 20 muestras mesolíticas tipadas para el mtDNA. En este caso el objetivo era comprobar la determinación del sexo realizada antropológicamente y excluir posibles contaminaciones procedentes de los

manipuladores. Para los tres individuos la amplificación del gen de la amelogénina permitió confirmar las hipótesis formuladas a partir del análisis morfológico de los esqueletos. Sin embargo, de los tres perfiles obtenidos, solo en un caso fue posible replicar los resultados de más de un marcador entre dos muestras diferentes procedentes del mismo individuo, apuntando a la elevada degradación del material genético y a la posible presencia de artefactos.

Entre las aplicaciones a muestras históricas, uno de los casos más conocidos es el de la identificación de la familia Romanov, zares de Rusia asesinados en 1918 por los bolcheviques. En 1991 se encontraron nueve esqueletos enterrados en Ekaterinburgo y se supuso que podría tratarse de los de la familia Romanov. Los análisis genéticos permitieron identificar los restos del zar, la zarina, y de tres de sus cinco hijos. Se supuso que los cuatro cuerpos restantes correspondían al médico y a tres de los sirvientes (Gill et al., 1994). En primer lugar, se llevó a cabo un estudio de relaciones de parentesco mediante la determinación molecular del sexo y el análisis de STRs nucleares. Para la identificación del zar y la zarina se empleó el mtDNA, comparándolo con descendientes vivos de ambos por línea materna. Posteriormente, en 2007, un grupo de arqueólogos encontró a unos 70 metros de la tumba de la familia Romanov los restos de otros dos individuos. El análisis genético reveló que se trataba de los otros dos hijos de los zares que no se habían encontrado en la primera tumba (Coble et al., 2009). Además, gracias a los avances tecnológicos, fue posible llevar a cabo un estudio genético exhaustivo que incluía más marcadores nucleares así como del cromosoma Y. Gracias a esta contribución fue posible identificar de manera irrefutable a toda la familia Romanov.

Otro caso interesante es el estudio de los putativos restos del rey Luis XVII a partir de una muestra de corazón (Jehaes et al., 2001). Debido al estado de preservación de la muestra no fue posible amplificar ADN autosómico, pero sí el gen de la amelogénina y el mtDNA. Los resultados apuntaron a que se trataba de un varón con una secuencia de mtDNA idéntica a la de sus putativos descendientes por línea materna. Sin embargo, debido a las características de este marcador, no fue posible discernir entre el hecho de que se tratara de Luis XII, hijo de María Antonieta y Luis XVI, u otro miembro de la familia a lo largo de la línea materna.

Un ejemplo de identificación de relaciones familiares en un contexto histórico es el estudio genético realizado sobre los restos de una familia noble del siglo XVI encontrados en la Iglesia de Santa Margarita en Reichersdorf, Alemania (Gerstenberger et al., 1999). Los datos genéticos se compararon con las inscripciones de las tumbas, revelando que a pesar de que supuestamente había ocho individuos masculinos enterrados, al menos dos eran mujeres. Fue posible además amplificar marcadores del cromosoma Y, que permitieron identificar un intercambio de esqueletos entre dos tumbas gracias a la comparación con descendientes vivos. También se averiguó que el hijo de uno de los individuos estudiados era ilegítimo o que no pertenecía a la familia enterrada en la iglesia, puesto que no compartía el mismo cromosoma Y con su supuesto progenitor. El estudio también incluía marcadores nucleares, gracias a los cuales fue posible identificar las relaciones familiares entre las dos mujeres, madre e hija. Además, el supuesto individuo no perteneciente a la familia resultó ser el padre biológico de esta última, confirmando que se trataba de un hijo ilegítimo.

Existen muchos otros estudios de este tipo, que suelen incluir el análisis de relaciones familiares así como intentos de identificación a través de sus descendientes. Por ejemplo, han sido objeto de estudio los putativos restos del evangelista Lucas (Vernesi et al., 2001), los del astrónomo Copérnico (Bogdanowicz et al., 2009) o del literato italiano del siglo XIV Francesco Petrarca (Caramelli et al., 2007), entre otros.

Sin embargo, uno de los casos que más ha llamado la atención del público es la reciente publicación de un estudio de relaciones familiares a partir de restos mucho más antiguos: la familia de Tutankamon (Hawass et al., 2010). Los autores analizaron 11 momias de hace unos 2.500 años, identificando relaciones familiares entre ellas y sus posibles causas de muerte. Gracias al estudio de marcadores nucleares y del cromosoma Y fue posible identificar a los padres del faraón Tutankamon. Asimismo se detectó molecularmente la presencia del *Plasmodium falciparum*, patógeno responsable del desarrollo de la malaria, en cuatro momias, incluida la de Tutankamon. Este estudio hace hincapié en cómo la sinergia de diferentes disciplinas

como la genética, la antropología y la arqueología, pueden permitir la realización de estudios multidisciplinares y ayudar a comprender eventos pasados.

OBJETIVOS

OBJETIVOS (ESPAÑOL)

La presente tesis doctoral se estructura en tres grandes bloques que se ven reflejados en todos sus apartados. Por tanto los objetivos están subdivididos en:

- Adecuación metodológica de las técnicas disponibles para el análisis genético de restos humanos antiguos.
- Estudio diacrónico-poblacional de diferentes grupos humanos que poblaron el noreste de la Península Ibérica en diferentes períodos prehistóricos.
- Establecimiento de relaciones familiares y determinación molecular del sexo en muestras históricas procedentes de enterramientos múltiples.

Metodología y validación de los resultados

1. Comparar diferentes metodologías disponibles en la literatura de ADN antiguo y forense para el aislamiento y la amplificación de ADN de muestras arqueológicas. Seleccionar aquellos protocolos más eficientes y que se ajustan a las características de las muestras disponibles y a las finalidades concretas del estudio.
2. Evaluar la posibilidad de recuperar ADN a partir de muestras antiguas en función de diferentes variables como son el estado de preservación, la procedencia, la antigüedad, las características morfológicas del tejido y el contenido de ADN en los extractos estimado mediante métodos espectrofotométricos.
3. Validar los resultados de ADN mitocondrial y nuclear obtenidos a partir de muestras históricas y prehistóricas mediante los pertinentes criterios de autenticidad, aceptados internacionalmente.

Estudio diacrónico-poblacional

4. Estimar la continuidad/discontinuidad genética entre los pobladores prehistóricos (neolíticos, calcolíticos, de la Edad del Bronce e íberos) y actuales del noreste de la Península Ibérica (Comunidad Valenciana, Cataluña y Aragón) y evaluar el impacto genético del Neolítico en esta misma área.

Relaciones familiares

5. Verificar las hipótesis de parentesco y de determinación del sexo formuladas a partir de otras disciplinas (arqueología, historiografía, genealogía, antropología) mediante el análisis genético de marcadores nucleares STR en un conjunto de muestras medievales procedentes de sepulturas múltiples (Iglesia de San Esteban de Cuéllar, Segovia).

AIMS and SCOPE

AIMS AND SCOPE (ENGLISH)

The aims and scope of this PhD thesis focuses on three topics, which are divided into the following sections:

- Methodological optimisation of the available techniques for the genetic analysis of ancient human remains.
- Diachronic population study of different human groups that inhabited the North Eastern Iberian Peninsula in different prehistoric periods.
- Establishment of kinship relationships and molecular sex determination in historical samples from multiple tombs.

Methodology and validation of the results

1. Comparison of different methods, available in ancient and forensic DNA literature, for the isolation and amplification of DNA from archaeological samples. Selection of the most efficient protocols considering the characteristics of the available samples and the specific purposes of the study.
2. Evaluation of the success of DNA recovery from ancient samples based on different variables such as sample preservation, origin, age, morphological characteristics of the tissue and DNA content in the extracts measured using spectrophotometric methods.
3. Validation of mitochondrial and nuclear DNA data from historic and prehistoric samples using international criteria of authenticity.

Diachronic-population study

4. Estimate the genetic continuity/discontinuity between prehistoric settlers (Neolithic, Chalcolithic, Bronze Age and Iberic culture) and current inhabitants of the North Eastern Iberian Peninsula (Catalonia, Aragón and Valencia regions) and evaluate the Neolithic genetic impact in this area.

Kinship relationships

5. Verify the hypotheses of kinship and sex determination determined by other disciplines (Archaeology, Historiography, Genealogy, Anthropology) through genetic analysis of nuclear STR markers in a medieval sample set from multiple tombs (Church of San Esteban, Cuéllar, Segovia).

ARTÍCULOS/ARTICLES

ARTÍCULOS / ARTICLES

Estudios metodológicos / Methodological studies

ARTÍCULO 1 / ARTICLE 1

Autores/Authors: Gamba C, Fernández E, López-Parra AM, y Arroyo-Pardo E.

Título/Title: **Statistical evaluation of pre-laboratory and laboratory factors that influence DNA recovery from archaeological material.**

Revista/Journal: Forensic Science International: Genetics Supplement Series.

Fecha de publicación/Publication Date: aceptado en septiembre de 2011, en prensa / accepted in September 2011, in press.

ARTÍCULO 2 / ARTICLE 2

Autores/Authors: Gamba C, Baeza C, Fernández E, Tirado M, López-Parra AM, y Arroyo-Pardo E.

Título/Title: Validation of the MiniFiler™ Kit in archaeological samples.

Revista/Journal: Forensic Science International: Genetics Supplement Series 2:17-18.

Fecha de publicación/Publication date: 2009.

Estudios diacrónicos-poblacionales / Diachronical-population studies

ARTÍCULO 3 / ARTICLE 3

Autores/Authors: Gamba C, Fernández E, Tirado M, Deguilloux M-F, Pemonge M-H, Utrilla P, Edo M, Molist M, Rasteiro R, y Chikhi L, Arroyo-Pardo E.

Título/Title: **Ancient DNA from an Early Neolithic Iberian population supports a pioneer colonisation by first farmers.**

Revista/Journal: Molecular Ecology.

Fecha de publicación/Publication date: aceptado en octubre de 2011, en prensa / accepted in October 2011, in press.

ARTÍCULO 4 / ARTICLE 4

Autores/Authors: Gamba C, Fernández E, Oliver A, Tirado M, Baeza C, López-Parra A, y Arroyo-Pardo E.

Título/Title: **Population genetics and DNA preservation in ancient human remains from Eastern Spain.**

Revista/Journal: Forensic Science International: Genetics Supplement Series 1:462-464.

Fecha de publicación/Publication date: 2008.

Estudio de relaciones familiares / Kinship Relationships study

ARTÍCULO 5 / ARTICLE 5

Autores/Authors: Gamba C, Fernández E, Tirado M, Pastor F, y Arroyo-Pardo E.

Título/Title: **Brief communication: Ancient nuclear DNA and kinship analysis: The case of a medieval burial in San Esteban Church in Cuéllar (Segovia, Central Spain).**

Revista/Journal: American Journal of Physical Anthropology 144:485-491.

Fecha de publicación/Publication date: 2011.

ARTÍCULO 1 / ARTICLE 1

Incluye un anexo / Appendix included

Statistical evaluation of pre-laboratory and laboratory factors that influence DNA recovery from archaeological material.

Gamba C, Fernández E, López-Parra AM, y Arroyo-Pardo E.

Forensic Science International: Genetics Supplement Series. En prensa.

RESUMEN (ESPAÑOL)

El objetivo de este trabajo consistía en la evaluación de diferentes factores que pueden influir en el éxito del análisis genético de muestras antiguas. Se evaluaron la eficiencia de obtención de resultados y la presencia de inhibidores de la PCR en función de las siguientes variables: protocolo de extracción empleado, edad de la muestra, localización del yacimiento arqueológico, tipo de muestra, color e integridad.

Estos análisis se llevaron a cabo sobre un conjunto de 89 muestras de diferentes procedencias y épocas (9.400-1.300 AP). El material genético fue extraído mediante cuatro protocolos diferentes y se amplificó la región hipervariable I del ADN mitocondrial.

Los resultados apuntan a que ni la edad de las muestras ni sus principales características macroscópicas (color, integridad, tejido) afectan a la obtención de resultados. La eficiencia en la obtención de resultados se ve fuertemente determinada por la procedencia de la muestra, resaltando la importancia del entorno de preservación en la obtención de información genética reproducible. También el método de extracción empleado resulta un factor determinante, puesto que influye en el éxito de los análisis a nivel de eficiencia de obtención de resultados y en la eliminación de moléculas inhibitoras. Los resultados obtenidos apuntan a que el protocolo de extracción basado en el empleo de sílica (Rohland y Hofreiter 2007) es el más eficiente para la extracción de ADN a partir de restos arqueológicos.



Contents lists available at ScienceDirect

Forensic Science International: Genetics Supplement Series

journal homepage: www.elsevier.com/locate/FSIGSS



Statistical evaluation of pre-laboratory and laboratory factors that influence DNA recovery from archaeological material

Cristina Gamba^{a,*}, Eva Fernández^{a,b}, Ana María López-Parra^a, Eduardo Arroyo-Pardo^a

^a Laboratorio de Genética Forense y Genética de Poblaciones, Universidad Complutense de Madrid, 28040 Madrid, Spain

^b Instituto de Arqueología e Paleociencias (UNL/UALg), 1069-061 Lisbon, Portugal

ARTICLE INFO

Article history:

Received 11 August 2011

Received in revised form 23 August 2011

Accepted 25 August 2011

Keywords:

Ancient DNA
Mitochondrial DNA
DNA amplification
PCR inhibition

ABSTRACT

We evaluated the influence of different factors over the efficiency of DNA amplification and PCR inhibition from 89 archaeological samples. We studied the effect of sample age and location together with some macroscopic features, like sample colour and fragmentation. Four different extraction methods were also tested and mtDNA HVRI was amplified and sequenced. Amplification and sequencing efficiencies were evaluated considering all the variables mentioned by a bivariate statistical analysis using SPSS 15.0 software. Results suggested that neither age nor macroscopic preservation had a direct effect on the recovery of endogenous DNA, while sample origin showed to influence significantly extraction and amplification successes. Moreover, the extraction method employed seemed to be a determining factor, influencing both inhibition and efficiency. An overall analysis of the results identified a silica-based protocol as the most efficient method for extracting DNA from archaeological material.

© 2011 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Several factors have been suggested to directly influence DNA recovery from archaeological samples, such as taphonomic conditions [1–3], climatic conditions [4,5], type of sample [1] and experimental procedures [6,7]. However, not all aspects that influence ancient DNA preservation and amplification have been fully understood. Here we shed some light on this question by analyzing some characteristics of the samples that are supposed to influence an ancient DNA (aDNA) successful analysis. Thus, the effects of age, location of the archaeological site and of several macroscopic features of the samples, like colour and fragmentation, were studied. In addition, four different extraction protocols were tested as well as their relationship with the success rate and the inhibition power.

2. Materials and methods

We studied 89 ancient bone and teeth samples coming from different locations and ages: Spain (21 sites, 81 samples, 7300–1300 BP), Syria (1 site, 4 samples, 9400–9200 BP) and Ecuador (1

site, 4 samples, 1400 BP). The external surface of samples was removed with a Sand-blasted. Samples were then irradiated with UV light and crushed in a freezer mill. DNA extraction was carried out with four different methods: (a) a silica-based protocol [8] (S-B); (b) a modified phenol/chloroform protocol [9] (P/C); (c) the extraction protocol of Yang et al. [10] (Y); (d) the QIAamp ADN Micro Kit (Qiagen[®]) (Q). Moreover, two variants of the S-B were tested: one with GuSCN and the other with NaCl. A total number of 116 extracts were obtained.

The quantitation of DNA in each extract was estimated by spectrophotometry using a NanoDrop ND-1000. A 244 bp fragment of the mtDNA HVRI (16126–16369) was amplified into two overlapping fragments using two primer sets [9] using the Multiplex PCR Kit (Qiagen[®]). Amplicons were sequenced using an ABI-Prism 310 (Applied Biosystems[®]) sequencer. The inhibition power of aDNA extracts was measured by adding 5 or 2.5 µl of each into PCR reactions (25 µl, primers T7 and SP6) with 50 ng of standard DNA (pGEM[®]-T Vector, Applied Biosystems[®]). Absence or decrease in band intensity in 2% agarose gels was interpreted as presence of inhibitors in aDNA extracts. Most criteria of authenticity were followed in sample analysis.

All the collected information was compared with the experimental results by a bivariate statistical analysis using the software SPSS 15.0. We evaluated the influence of different factors (age, location, extraction method, macroscopic preservation and type of sample) over positive amplifications (*Amplification Success*), the replication of results (*Replication Efficiency*) and the PCR inhibition (*Inhibition Power*).

* Corresponding author at: Universidad Complutense de Madrid, Departamento de Toxicología y Legislación Sanitaria, Facultad de Medicina, Pabellón 7 - 4^a planta, Avenida Complutense s/n, 28040 Madrid, Spain. Tel.: +34 913941576; fax: +34 913941576.

E-mail address: cristinagamba@med.ucm.es (C. Gamba).

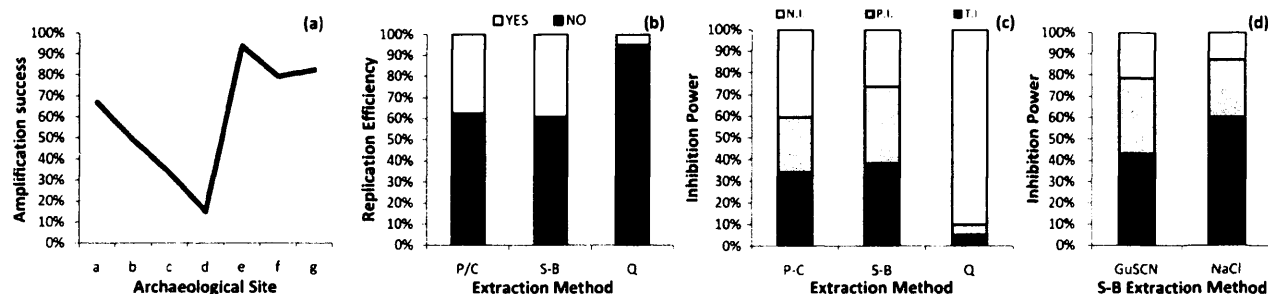


Fig. 1. (a) Amplification efficiency vs archaeological site. (b) Replication efficiency vs extraction method. (c) Extracts inhibition power vs extraction method (N.I. = no inhibition, P.I. = partial inhibition, T.I. = total inhibition). (d) Extracts inhibition power vs the silica-based extraction method.

3. Results and discussion

The age, the spectrophotometric estimation of DNA concentration, the macroscopic preservation and the type of sample (tooth or bone) did not show to have a direct influence neither on the *Amplification Success*, nor on the *Replication Efficiency* or the *Inhibition Power* (data not shown). However, statistical analysis showed that the *Amplification Success* depended on the location of the archaeological site (see panel a of Fig. 1). The extraction method influenced the possibility of obtaining reproducible results in two independent experiments. Specifically, the *Q method* demonstrated to have very low *Replication Efficiency*, on the contrary, the *P/C* and the *S-B protocols* showed a similar efficiency (see panel b of Fig. 1). The *Y method* did not give any result and was excluded from the statistical analysis. We showed that the *Inhibition Power* of the DNA extracts is also related to the extraction protocol employed. The *Q method* seemed to be very efficient in eliminating inhibitors, but the extracts obtained by this protocol did not give reproducible results. Even though the *S-B protocol* was globally less efficient than the *P/C* in eliminating inhibitors (see panel c of Fig. 1), both efficiencies were similar when using GuSCN in the former (see panel d of Fig. 1).

4. Conclusions

The results obtained indicate that ancient DNA preservation mainly depends on the depositional environment rather than on sample age. This is in agreement with previous observations [3]. However, sample macroscopic features, usually employed as a criteria for sampling selection, seem not to influence the final result of the analysis. The extraction method was the most influencing variable on both inhibition and aDNA recovery efficiency. An overall analysis of results pointed out the *S-B protocol* with GuSCN as the most efficient protocol, which is faster and less toxic than the *P/C method*.

Funding

This work was supported by CGL2006-07828/BOS and CGL2009-07959 (Spanish Government research projects) and by

the Dávalos-Fletcher Foundation (Castellón, Spain). Human resources were funded by an FPU grant (Spanish Government) for C.G. and a MICINN researcher contract “Juan de la Cierva” for E.F. (partially supported by the European Social Fund).

Conflict of interest

None.

Acknowledgements

We thank all the archaeologists and anthropologists who provided samples for these analyses and collaborated with this work: Miquel Molist, María del Carmen Molestina, Ana Blasco, Javier Jiménez, José María López and Domingo Martínez.

References

- [1] C.M. Nielsen-Marsh, R.E.M. Hedges, Patterns of diagenesis in bone I: the effects of site environments, *Journal of Archaeological Science* 27 (December (12)) (2000) 1139–1150.
- [2] J. Burger, S. Hummel, B. Hermann, W. Henke, DNA preservation: a microsatellite-DNA study on ancient skeletal remains, *Electrophoresis* 20 (June (8)) (1999) 1722–1728.
- [3] T. de Torres, J.E. Ortiz, F.J. Llamas, L. Canoira, R. Julia, M.J. Garcia-Martinez, Bear dentine aspartic acid racemization analysis: a proxy for the dating of Pleistocene cave infills, *Archaeometry* 44 (3) (2002 Ago) 417–426.
- [4] R. Bollongino, A. Tresset, J.-D. Vigne, Environment and excavation: pre-lab impacts on ancient DNA analyses, *Comptes Rendus Palevol* 7 (April (2–3)) (2008) 91–98.
- [5] R.E.M. Hedges, A.R. Millard, Bones and groundwater: towards the modelling of diagenetic processes, *Journal of Archaeological Science* 22 (March (2)) (1995) 155–164.
- [6] C. Anderung, P. Persson, A. Bouwman, R. Elburg, A. Götherström, Fishing for ancient DNA, *Forensic Science International: Genetics* 2 (March (2)) (2008) 104–107.
- [7] N. Rohland, M. Hofreiter, Comparison and optimization of ancient DNA extraction, *BioTechniques* 42 (March (3)) (2007) 343–352.
- [8] N. Rohland, M. Hofreiter, Ancient DNA extraction from bones and teeth, *Nature Protocols* 2 (July (7)) (2007) 1756–1762.
- [9] E. Fernández, Polimorfismos de DNA mitocondrial en poblaciones antiguas de la cuenca mediterránea, PhD Thesis, 2005.
- [10] D.Y. Yang, B. Eng, J.S. Wayne, J.C. Dudar, S.R. Saunders, Technical note: improved DNA extraction from ancient bones using silica-based spin columns, *American Journal of Physical Anthropology* 105 (April (4)) (1998) 539–543.

ANEXO - ARTÍCULO 1

Tabla A1. Resumen de las diferencias significativas (+) y no significativas (-) para las variables "Yacimiento", "Provincia" y "Época".

	Yacimiento	Provincia	Época
Amplificaciones positiva	+	+	+
Secuencias no contaminadas	+	+	+
Replicación de los resultados	-	-	-
Inhibición de la PCR	+	+	-

Tabla A2. Resumen de las diferencias significativas (+), casi significativas (+/-) y no significativas (-) para las variables "Pieza", "Integridad" y "Color".

Eficiencia	Pieza	Integridad	Color
Amplificaciones positiva	+	-	+/-
Secuencias no contaminadas	+	+	-
Replicación de los resultados	-	-	-
Inhibición	-	-	-



APPENDIX - ARTICLE 1

Table A1. Summary of statistically significant (+) and not significant (-) differences for the variables "Site", "Province" and "Age".

	Site	Province	Age
Positive amplifications	+	+	+
Sequences without contamination	+	+	+
Replication of results	-	-	-
PCR inhibition	+	+	-

Table A2. Summary of statistically significant (+), marginally significant (+/-) and not significant (-) differences for the variables "Type of sample" (tooth or bone), "Integrity" and "Colour".

	Type of sample	Integrity	Colour
Positive amplifications	+	-	+/-
Sequences without contamination	+	+	-
Replication of results	-	-	-
PCR inhibition	-	-	-

ARTÍCULO 2 / ARTICLE 2

Validation of the MiniFiler™ Kit in archaeological samples.

Gamba C, Baeza C, Fernández E, Tirado M, López-Parra AM, y Arroyo-Pardo E.

Forensic Science International: Genetics Supplement Series 2:17-18. 2009.

RESUMEN (ESPAÑOL)

En este trabajo se presentan los resultados de la evaluación de la utilidad de un kit para la amplificación de STRs nucleares desarrollado para el campo forense en el estudio de muestras humanas antiguas. Se trata del kit *MiniFiler* (*Applied Biosystems*), especialmente diseñado en 2008 para la amplificación de muestras degradadas.

Se analizó un amplio espectro de muestras, todas procedentes de la Península Ibérica pero de diferentes épocas: 6 muestras neolíticas, 14 de la Edad del Bronce, 15 medievales y 5 del siglo XVIII. Fue posible demostrar la utilidad del kit testado en muestras bien preservadas, como el conjunto de muestras esqueléticas y momificadas medievales. Sin embargo, en el caso de muestras más antiguas únicamente se pudieron recuperar perfiles muy parciales y no reproducibles en diferentes amplificaciones. Se trata de un resultado esperable dado que en muestras de tal antigüedad resulta complicado amplificar marcadores nucleares debido a la elevada degradación del ADN. Por otro lado, resulta sorprendente la ausencia de resultados en el conjunto de muestras más recientes, del siglo XVIII. Este resultado apunta a la degradación acelerada de estas muestras, probablemente debida a la elevada humedad del entorno de preservación y a la absorción de sustancias inhibidoras del suelo.

En su conjunto, estos resultados evidencian la importancia del ambiente tafonómico para la buena preservación de las muestras antiguas y del material genético que contienen.

Research article

Validation of the *MiniFiler*TM Kit in archaeological samplesCristina Gamba ^{*}, Carlos Baeza, Eva Fernández, Mirian Tirado,
Ana María López-Parra, Eduardo Arroyo-Pardo

Toxicology and Health Legislation Department, Faculty of Medicine, Complutense University of Madrid, Spain

Received 29 July 2009; accepted 1 August 2009

Abstract

We evaluate the usefulness of *MiniFiler*TM Kit in the field of ancient DNA. A set of samples belonging to different locations from Iberian Peninsula, with ages ranging from Neolithic to XVII century, was tested. Results could be replicated in only one burial site, probably due to the taphonomic conditions. Other cases could only produce partial or none genetic profiles.

© 2009 Elsevier Ireland Ltd. All rights reserved.

Keywords: Ancient DNA; MiniSTRs; Human nuclear DNA; Spain

1. Introduction and aims

The forensic *MiniFiler*TM Kit has been designed to amplify nuclear DNA from critical samples. Amplification success is incremented through three different strategies: increasing the sensibility, decreasing the length of the amplicons and bypassing inhibition problems [2]. We pretend to evaluate the usefulness of the forensic *MiniFiler*TM Kit (Applied Biosystems) for ancient DNA (aDNA). In this study, we tested the *MiniFiler*TM Kit within a wide range of ancient Spanish samples, from different ages and locations, all of them preserved under different taphonomic conditions.

In the archaeogenetic field, obtaining nuclear DNA information could lead to confirm or reject archaeological hypothesis, to establish relationships between individuals belonging to the same burial site, to determine the sex or to identify different samples belonging to the same individual.

2. Materials and methods

We studied 6 Neolithic samples from the “Can Sadurní” Cave (Barcelona), 14 Bronze Age samples from 4 different archaeological sites of the Castilla y León Region (North West Spain), 15 Medieval samples from “El Cuellar” (Segovia,

Central Spain) and 5 samples of the XVIIIth century from Murcia (South East Spain).

DNA cleaning, grinding, extractions and amplifications were performed in an ancient DNA specialised laboratory, following the recommended criteria of authenticity [3].

DNA extractions were performed over bones and teeth through a silica-based method by Rohland and Hofreiter, 2007 [4]. Amplifications were carried out according to the manufacturer's recommendations. The *MiniFiler*TM Kit is designed to amplify 8 autosomal STR loci (D13S317, D7S820, D2S1338, D21S11, D16S539, D18S51, CSF1PO, FGA) and the gender-identification region Amelogenin. Amplified STRs markers were separated on a 3730 DNA Analyzer (Applied Biosystems).

In order to validate results, a minimum of 2 independent *MiniFiler*TM PCRs was performed. For samples from “El Cuellar” it was possible to increase the reliability of results by sampling at list 2 teeth or bones per individual. More precisely, the 15 samples from “El Cuellar” belong to 7 different individuals: 5 adults and 2 infants.

Kinship relationship between samples which yielded partial or complete genetic profiles was estimated using the *Familias 1.81* software.

3. Results and discussion

Results are summarised in Table 1 and Fig. 1.

We obtained very satisfactory results with the well-preserved – and in some cases mummified – samples from

^{*} Corresponding author at: Dpto. Toxicología y Legislación Sanitaria, Facultad de Medicina, Universidad Complutense de Madrid, 28040 Madrid, Spain. Tel.: +34 91 394 1576; fax: +34 91 394 1576.

E-mail address: cristinagamba@med.ucm.es (C. Gamba).

Table 1
Samples studied (origin, number of samples and individuals) and results obtained in terms of reproducibility.

Samples origin	No. individuals	No. samples	% replicated results	% not replicated results	% no results
El Cuellar	7	14	82.54	3.17	14.29
Cogotas Culture	14	14	3.97	16.67	79.37
Can Sadurní	8	8	2.78	31.94	65.28
Murcia	5	5	0.00	0.00	100.00
	34	41	22.32	12.95	64.73
Total			Average		

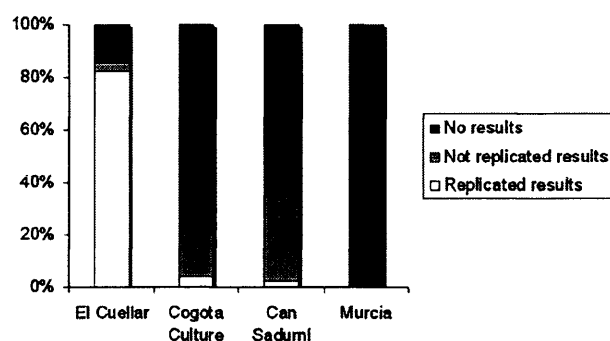


Fig. 1. Histograms representing the relative percentages of reproducible results in correspondence to the archaeological site studied.

“El Cuellar” (Middle Ages), buried in sarcophagus. It was possible to obtain the complete genetic profile from all the adult samples as well as partial profiles from 2 newborn infants. Amelogenin marker allowed us to establish the gender for all the 7 samples. We compared the gender results with the historical, archaeological and anthropological data. The individual buried in a sarcophagus with a female name inscription was classified as masculine skeleton through anthropometric criteria. However, genetic analysis gave the opposite result. A double replication per each sample supported the grave inscription. Globally, for this site, it was possible to obtain a high percentage of replicated results: 82.54% of replication per marker and per individual (see Table 1). Genetic information obtained was practically complete and allowed us to estimate paternity and brotherhood indexes between some of the individuals. Statistical analysis of the kinship relationship between 2 individuals, a mummy and a skeleton, from “El Cuellar” gave a relatively high statistical probability (99.98%). This result supports a previous archaeological hypothesis and demonstrates the importance and power of interdisciplinary studies in archaeogenetics.

Regarding the most recent samples from Murcia, we did not obtain positive results, most likely due to the burial conditions, with a high percentage of humidity, which increases the DNA degradation [1].

Neolithic and Bronze Age samples yielded only very partial profiles, with a low rate of replication in amplifications. In both cases, we studied apparently well-preserved teeth. However, due to their older age, it is important to take into account a longer exposure to the taphonomic environment and subsequently a higher DNA degradation.

4. Conclusions

This study demonstrates the usefulness of the *MiniFiler*TM Kit in the field of ancient DNA, specifically in the case of well-preserved historical samples, such as the case of “El Cuellar” samples. Results from prehistoric materials (Cogotas Culture and Can Sadurní Cave) did not show neither sufficient reproducibility nor markers enough of the total set comprised in *MiniFiler*TM Kit.

Role of funding

This study was funded by a MICINN project (CGL2006-07828/BOS) and a CAM-UCM project (CCG08-UCM/BIO-3938).

Conflict of interest

None.

Acknowledgements

Thanks to archaeologists and anthropologists who provide samples for this study: Francisco Pastor (El Cuellar, U. Valladolid), Àngel Esparza (Cogotas Culture, U. Salamanca), Anna Blasco (Can Sadurní, U. Barcelona) and José Antonio Sánchez (Murcia, UCM).

References

- [1] T. Lindahl, Instability and decay of the primary structure of DNA, *Nature* 362 (1993) 709–715.
- [2] J. Mulero, et al., Development and validation of the AmpF^{STR} MiniFilerTM PCR amplification kit: a MiniSTR multiplex for the analysis of degraded and or PCR inhibited DNA, *J. Forensic Sci.* 53 (4) (2008) 838–852.
- [3] S. Pääbo, et al., Genetic analyses from ancient DNA, *Annu. Rev. Genet.* 38 (2004) 645–679.
- [4] N. Rohland, M. Hofreiter, Ancient DNA extraction from bones and teeth, *Nat. Protoc.* 2 (7) (2007) 1756–1762.

ARTÍCULO 3 / ARTICLE 3

Incluye Material Suplementario / Supplementary Material included

Ancient DNA from an Early Neolithic Iberian population supports a pioneer colonisation by first farmers.

Gamba C, Fernández E, Tirado M, Deguilloux M-F, Pemonge M-H, Utrilla P, Edo M, Molist M, Rasteiro R, y Chikhi L, Arroyo-Pardo E.

Molecular Ecology. En prensa.

RESUMEN (ESPAÑOL)

El origen de la población europea ha sido objeto de estudio de diferentes disciplinas, como la arqueología, la antropología y la genética. Uno de los temas más debatidos es la contribución demográfica que pudo acompañar a la difusión del neolítico desde Próximo Oriente hace 10.000 años.

El objetivo de este trabajo era proporcionar nuevos datos acerca de la transición al Neolítico en el noreste de la Península Ibérica mediante el análisis de la región Hipervariable I del ADN mitocondrial de los primeros pobladores neolíticos de la región.

Se obtuvieron resultados validados para 13 individuos procedentes de tres sitios arqueológicos: Can Sadurní (Neolítico Antiguo Cardial), Chaves (Neolítico Antiguo Cardial) y Sant Pau del Camp (Neolítico Antiguo Post-Cardial). Estos resultados se compararon con aquellos de la misma región disponibles en la literatura: una población Neolítica Media y los habitantes actuales.

La población neolítica estudiada estaba caracterizada por la presencia de haplogrupos mitocondriales (N*y X1) poco comunes en la población europea actual. Para explicar los resultados obtenidos se emplearon métodos computacionales basados en simulaciones bayesianas, que destacaron una elevada diferenciación genética entre la población neolítica antigua y las poblaciones de Neolítico Medio y moderna).

Los datos aquí presentados son compatibles con la llegada de pequeños grupos desde Próximo Oriente a principios del Neolítico. También sugieren que durante este período la deriva genética jugó un papel importante y que el modelo que mejor se ajusta a los resultados es una colonización pionera por parte de estos primeros grupos neolíticos en el noreste de la Península Ibérica.

Ancient DNA from an Early Neolithic Iberian population supports a pioneer colonization by first farmers

C. GAMBA,* E. FERNÁNDEZ,**† M. TIRADO,* M. F. DEGUILLLOUX,‡ M. H. PEMONGE,‡ P. UTRILLA,§ M. EDO,¶ M. MOLIST,** R. RASTEIRO,†† L. CHIKHI††‡§§ and E. ARROYO-PARDO*

*Laboratorio de Genética Forense y Genética de Poblaciones, Facultad de Medicina, Pabellón 7, 4ª Planta, Universidad Complutense de Madrid, Avenida Complutense s/n, 28040 Madrid, Spain, †Instituto de Arqueologia e Paleociências, Universidade do Algarve, 8005-139 Faro, Portugal, ‡UMR 5199 PACEA, Laboratoire d'Anthropologie Des Populations Passées et Présentes, Université Bordeaux 1, 33405 Talence cedex, France, §Departamento de Ciencias de la Antigüedad, Universidad de Zaragoza, 50009 Zaragoza, Spain, ¶Departament de Prehistòria, Història Antiga i Arqueologia, Universitat de Barcelona, 08032 Barcelona, Spain, **Departamento de Prehistoria, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain, ††Instituto Gulbenkian de Ciência, P-2780-156 Oeiras, Portugal, ‡‡CNRS, Université Paul Sabatier, ENFA; UMR 5174 EDB (Laboratoire Evolution & Diversité Biologique); 118 route de Narbonne, F-31062 Toulouse, France, §§Université de Toulouse; UMR 5174 EDB, F-31062 Toulouse, France

Abstract

The Neolithic transition has been widely debated particularly regarding the extent to which this revolution implied a demographic expansion from the Near East. We attempted to shed some light on this process in northeastern Iberia by combining ancient DNA (aDNA) data from Early Neolithic settlers and published DNA data from Middle Neolithic and modern samples from the same region. We successfully extracted and amplified mitochondrial DNA from 13 human specimens, found at three archaeological sites dated back to the Cardial culture in the Early Neolithic (Can Sadurní and Chaves) and to the Late Early Neolithic (Sant Pau del Camp). We found that haplogroups with a low frequency in modern populations—N* and X1—are found at higher frequencies in our Early Neolithic population (~31%). Genetic differentiation between Early and Middle Neolithic populations was significant ($F_{ST} \sim 0.13$, $P < 10^{-5}$), suggesting that genetic drift played an important role at this time. To improve our understanding of the Neolithic demographic processes, we used a Bayesian coalescence-based simulation approach to identify the most likely of three demographic scenarios that might explain the genetic data. The three scenarios were chosen to reflect archaeological knowledge and previous genetic studies using similar inferential approaches. We found that models that ignore population structure, as previously used in aDNA studies, are unlikely to explain the data. Our results are compatible with a pioneer colonization of northeastern Iberia at the Early Neolithic characterized by the arrival of small genetically distinctive groups, showing cultural and genetic connections with the Near East.

Keywords: ancient DNA, Iberian Peninsula, mitochondrial DNA, Neolithic

Received 8 July 2011; revision received 5 October 2011; accepted 11 October 2011

Introduction

The Neolithic transition that transformed Europe arose in the Near East more than 10 000 BP and spread

towards Western and Central Europe during an expansion that lasted several millennia (Price 2000). It represents a major if not the most important economic revolution to take place in Western and Central Europe. Foraging was replaced by agriculture and animal farming, eliciting a simultaneous demographic response characterized by a sharp increase in birth rate (Bocquet-Appel & Bar-Yosef

Correspondence: Cristina Gamba, Fax: +34 913941576; E-mail: cristinagamba@med.ucm.es

2008). The last 50 years have witnessed a major scientific controversy over how this spread took place, basically focused on its demographic or cultural nature, usually defined by two extreme models of *cultural diffusion* (CDM) and *demic diffusion* (DDM) (Childe 1964; Zvelebil 2001). Intermediate models, such as *Infiltration*, *Leapfrog colonization*, *Folk migration*, *Frontier mobility*, *Elite dominance* or *Contact*, have been also suggested (Zvelebil 2001). In the Iberian Peninsula, archaeological data suggest a dual model (coastal DDM and inland CDM) (Bernabeu 1997; Fernández & Gómez 2009) and radiocarbon dates support a rapid spread of Neolithic culture in a framework of a maritime pioneer colonization (Zilhão 2001), which suppose the arrival of small Neolithic groups to coastal areas.

To quantify the relative genetic contribution of Near Eastern Neolithic populations to the European gene pool, modern genetic variability has been explored with contradictory results depending on the type of markers studied and/or the approaches involved in data analyses (Richards 2003; Dupanloup *et al.* 2004; Barbujani & Chikhi 2006; Soares *et al.* 2010). Whereas phylogeographic studies using Y chromosome and mitochondrial DNA suggested limited contributions (around 20–30%) and concluded in favour of the CDM (Richards *et al.* 2000; Semino *et al.* 2000), simulation analyses performed with some of these data provided a much wider range of possible contributions, perhaps as much as 80%, depending on the populations analysed (Chikhi *et al.* 2002), and favoured a DDM.

Ancient DNA (aDNA) analyses have made a step forward in the last few years. Although this approach has some limitations, such as a restricted number of samples, it provides valuable first-hand information about human ancestry. The Neolithic transition in Central Europe has been explored using aDNA and up to 42 Neolithic skeletons associated with the Linear Pottery Culture [LBK, *Linearbandkeramik*, 5500–4900 cal. before the common era (BCE)] were successfully typed (Haak *et al.* 2005, 2010). Moreover, the Mesolithic background in the same region was also studied (Bramanti *et al.* 2009). The LBK Neolithic population showed high frequency of haplogroup N1a (~25%), a haplogroup that is very rare in the modern-day European population (0.2%). This result was interpreted as suggesting a Palaeolithic ancestry for modern Europeans (Haak *et al.* 2005). This haplogroup was then proposed to be a genetic signature of the Neolithic spread in Central Europe, which was supported by its absence in the Mesolithic population of the surrounding areas (Bramanti *et al.* 2009). Haplogroup N1a was later detected in one Neolithic individual from a French Megalithic burial, suggesting that these interpretations might be incorrect (Deguilloux *et al.* 2011). The dispersion of this

lineage through pioneer Neolithic groups to western France was proposed by these authors. Recently, Haak *et al.* (2010) analysed all LBK mitochondrial lineages available at the time and showed that they had affinities with modern populations from the Near East. They concluded that it favoured the DDM with an important genetic contribution during the Neolithic spread towards Europe, hence contradicting the previous study of Haak *et al.* (2005).

While many 'phylogeographical' studies have tended to use network-based 'methods' to 'reconstruct' the demographic history of human populations, there is now an increasing recognition that evolutionary factors (e.g. mutation rate and genetic drift) are intrinsically noisy, that haplogroups cannot be identified to populations (Barbujani *et al.* 1998) or cultures and that model-based approaches are necessary to make progress in the statistical analyses of complex population scenarios (Beaumont *et al.* 2010), such as the Neolithic contribution to the European gene pool.

The colonization of Europe is thought to have taken place following two main routes, namely the Central European and the Mediterranean route of Neolithic spread. In contrast to the former, the Mediterranean route is linked to the Cardial pottery complex instead of the LBK and has been poorly studied. In this framework, Iberia represents an interesting case study, as it is located at the westernmost edge of the Neolithic expansion. The only Iberian Neolithic aDNA study in that region was carried out on Middle Neolithic (MN) samples (3500–3000 cal. BCE) from northeastern Iberia (Sampietro *et al.* 2007). Haplogroup frequencies were very similar to those of present-day populations and were interpreted as favouring a genetic continuity between Neolithic and modern-day Iberian populations. In the light of these divergent results between the Iberian Peninsula and Central Europe (because only the Haak *et al.* 2005 study had then been published), the authors assumed that two different mechanisms of Neolithic diffusion were involved: *cultural diffusion* in Central Europe and *demic diffusion* in the Mediterranean (Sampietro *et al.* 2007). However, some bias could have been introduced considering the MN data (3500–3000 cal. BCE) as representative of the first Neolithic settlers of the region (around 5500 BCE) owing to the age gap between them.

Here, we address this issue by presenting the first Early Neolithic (EN) (Cardial and post-Cardial pottery cultures) aDNA data from northeastern Iberian specimens. Data from these first Neolithic settlers were compared with later inhabitants of the same region in a diachronic context using different simulation scenarios.

Our results show that the simple panmictic models assumed by previous aDNA studies (with the exception

of Haak *et al.* 2010; R. Rasteiro and L. Chikhi, submitted) are rejected. We show that the genetic data support structured models that are compatible with a pioneer colonization by the first Neolithic groups in northeastern Iberia.

Materials and methods

Sample information

Forty-nine samples from a minimum of 22 individuals found in three Neolithic sites of northeastern Iberia were studied: Can Sadurní ($N = 7$, Barcelona province, 5475–5305 cal. BCE, Blasco *et al.* 2005), Chaves ($N = 3$, Huesca province, 5329–4999 cal. BCE, Utrilla *et al.* 2008) and Sant Pau del Camp ($N = 12$, Barcelona province, 4250–3700 cal. BCE, Molist *et al.* 2008) (see sites location in Fig. 1 and sample description in Table S1, Supporting information). Can Sadurní and Chaves samples are dated back to the EN and are associated with the Cardial culture. Sant Pau del Camp samples are slightly more recent and are dated back to Epicardial or Late EN. Can Sadurní archaeological site is a cave located at the Garraf mountains, 450 m above sea level and about 25 km west of Barcelona (northeastern Spain). Excavations started in 1978 and are still in progress. Twenty-eight different levels have been identified, ranging from Epipalaeolithic (10 840–10 410 cal. BCE) to the Roman period. In layer 18, 80% of impressed pottery was decorated with Cardium shells. Inside one of these Cardial potteries, a conglomerate of seeds was found and radio-

carbon-dated (5475–5305 cal. BCE, Blasco *et al.* 2005). Human skeletons were found in graves but not in anatomical connection owing to the fall of the entrance of the cave. Twenty-four loose teeth were selected following external preservation criteria. Dental study identified a minimum of 5 individuals (Blasco *et al.* 2005), but by comparing this information with archaeological and genetic results presented here, we were able to establish a minimum of seven individuals. The Chaves archaeological site is also a cave, located in Bastarás, Huesca province (northeastern Spain). Excavations took place from 1984 to 1992, identifying a long-time occupation of the site, spanning from the Palaeolithic (Solutrean, 19 390–20 010 cal. BCE, Montes & Utrilla 2008) to the Bronze Age. The five teeth studied belonged to three individuals, which were found in anatomical connection and come from the Cardial Neolithic level, where radiocarbon datings were directly performed on human bones (5329–4999 cal. BCE, Utrilla *et al.* 2008). Sant Pau del Camp is an open-air site found in the city of Barcelona. Excavations performed at the church of the same name revealed different occupation phases, ranging from the Early Neolithic, associated with Cardial pottery, to the Roman period. The 21 studied samples came from 12 individuals and were found in the Post-Cardial level, where human bones in anatomical connection were radiocarbon-dated (4250–3700 cal. BCE, Molist *et al.* 2008).

Entire teeth samples, without external fissures or caries, were selected by the Population Genetics and Forensic Genetic Laboratory staff, with the exception of two bones from the 1CH0102 specimen (Chaves site), for which dental samples were not available. Whenever possible, at least two teeth or bones per individual were selected. Further information about samples is provided in Table S1 (Supporting information).

Ancient DNA analyses

Criteria of authenticity. Genetic analyses were carried out in specialized aDNA laboratories located in the same building but with physical separation of Pre-polymerase chain reaction (PCR), PCR and Post-PCR procedures. Sample cleaning, grinding and extraction were performed in the Pre-PCR laboratory. PCRs were set up in the PCR laboratory and then amplified in a distant-separated room in which no DNA analyses were performed. Agarose gel analysis, PCR purification, cloning and sequencing were performed in a post-PCR laboratory. Pre-PCR and PCR laboratories were UV-irradiated before and after each experiment. Workbenches and laboratory equipment were regularly cleaned with 70% bleach to reduce carry-over contamination. Staff access to aDNA laboratories was restricted



Fig. 1 Archaeological sites location and population groups. The three population groups analysed here: (i) Early Neolithic samples (present study, Can Sadurní, Chaves and Sant Pau del Camp sites, white spots), (ii) Middle Neolithic samples (Sampietro *et al.* 2007; Camí de Can Grau site, black spot) and (iii) Modern samples from the same region (García *et al.* 2011, Catalonia and Aragón, shady).

to three people. Moreover, in this case, all experimental analyses were carried out by a single researcher (CG) to reduce exogenous DNA contamination. Sample cleaning and grinding, DNA extraction and PCR amplifications were carried out wearing disposable laboratory coveralls, masks, caps, glasses, shoe covers and gloves. All reagents and consumables employed were DNase and RNase free. All procedures were carried out in a laminar flow cabinet previously cleaned with bleach and UV-irradiated. Laboratory contamination was monitored with extraction blanks (one each seven samples), and at least three PCR-negative controls were included every seven samples. Only samples without any contamination in all stages were considered for further analyses (see Results).

The silica-based extraction method (Rohland & Hofreiter 2007) together with the Multiplex PCR kit (Qiagen) used to amplify aDNA in multiplex reactions proved to be a powerful tool for increasing aDNA yield and typing. Because this amplification kit has been shown to be very sensitive to contamination (data not shown), we designed a set of experiments in which noncontaminated extraction blanks (EBs), previously amplified at least twice, were re-amplified repeatedly to detect possible contamination. The overall results showed a 7% random contamination of EBs, neither related to staff nor to reagents but possibly attributed to random carry-over from floating molecules. Despite this evidence about difficulties in complete elimination of exogenous contamination at these phases, our data showed a low incidence of this problem and almost no contamination reproducibility in the same PCR set, so there was no chance of confusion with endogenous DNA data. To identify the potential sources of contamination, genetic profiles were recovered from all people involved in sample manipulation, including laboratory staff, anthropologist and archaeologists (Table S5, Supporting information). Sampling was performed by collecting blood spots on Whatman filter paper; samples were extracted with the QIAamp DNA Mini kit (Qiagen) and amplified with primers 1A and 1B under standard conditions (Table S6, Supporting information).

Results' reproducibility was assessed by setting up independent extractions and amplifications from the same skeleton. When different bones or teeth from the same individual were not available, two portions of the same tooth/bone were analysed in parallel. At least two amplifications of each mitochondrial DNA (mtDNA) region were performed on all aDNA extracts. Only consistent results among extractions and amplifications were considered. Moreover, five of 14 individuals typed were replicated by CG in an independent aDNA laboratory (LAPP, PACEA, Bordeaux 1 University—CNRS, France), using similar methodology and facilities as

those described above. Consistent independent amplifications were cloned (Table S7, Supporting information), and endogenous DNA was identified taking care to follow reproducibility criteria. Staff contamination was identified by comparing the results with staff genetic profiles, and carry-over contamination by comparing the results from the same round of cleaning, DNA extraction or amplification. Moreover, frequencies of transitions (type I and II) and transversions derived from molecular damage were calculated. Authenticity of results is supported by the excess of type II transitions (CG → TA, 68%) over type I transitions (TA → CG, 28%) and transversions (4%), as already suggested in the literature (Brotherton *et al.* 2007; Gilbert *et al.* 2007). Haplogroup single nucleotide polymorphism (SNP) typing results were in concordance with HVR-I information, supporting the authenticity of the recovered haplotypes.

Sample cleaning and grinding. Samples were cleaned using a Sand Blaster (Dentalfarm Base 1 Plus), which allows the removal of about a millimetre of the bone/tooth surface using aluminium oxide powder under pressure. The aim of this procedure is to clean the sample and remove contaminant DNA molecules from its outer surface. Samples were then irradiated with UV light for about 30 min per side in a laminar flow cabinet and transferred to sterile grinding vials. Grinding was performed in a Freezer Mill (SPEX Model 6700) filled with liquid nitrogen. The resulting powder was stored at -20 °C until DNA extraction was performed.

DNA extraction. DNA was extracted from approximately 500 mg of powdered sample following the protocol published by Rohland & Hofreiter (2007). In this protocol, DNA is absorbed to silica in the presence of high concentrations of a chaotropic salt, guanidinium thiocyanate (GuSCN).

Mitochondrial DNA amplification. Two overlapping fragments of the hypervariable region I (HVR-I) of the mtDNA were amplified by PCR. The primer pairs used (see Table S6 for primer sequences, Supporting information) allowed typing 294 bp (bases pairs) of the HVR-I (positions 16106–16399). PCRs were set up using the Qiagen Multiplex PCR kit in a final volume of 25 µL (12.5 µL Qiagen Multiplex PCR Master Mix, 9 µL RNase-Free Water, primers at 2.5 mM and 3 µL DNA extract). Hot-start amplifications were carried out in a Multigene II Personal Thermal Cycler (Labnet International, Inc.), consisting of an initial denaturation at 95 °C for 15 min, 40 cycles of 30 s at 94 °C, 90 s at 55–59 °C, 90 s at 72 °C, and a 10-min final extension at

72 °C. Amplifications were checked in 2% agarose gels, and positive amplifications were purified using the QIAquick PCR Purification kit (Qiagen), obtaining a final volume of 30 µL.

Sequencing and cloning. Purified PCR products were directly sequenced in ABI Prism 310 Genetic Analyzer (Applied Biosystems, Life Technologies) or sent to the C.A.I. laboratory (Centro de Apoyo a la Investigación, Research Support Centre, Universidad Autónoma de Madrid, Cantoblanco, Madrid, Spain) and run in ABI Prism 3700 Genetic Analyzer (Applied Biosystems, Life Technologies). Consistent amplifications were cloned using the pGEM-T Easy Vector System (Promega) or TOPO TA Cloning kit (Invitrogen) following the manufacturer's instructions. Twenty colonies with insert were grown in a different plate. Presence of DNA insert was then double-checked by colony PCR using vector primers SP6 and T7. PCRs were performed in a final volume of 10 µL under standard conditions and reagents (Bio-tools B&M LABs, S.A.). Amplifications were carried out in an Eppendorf Mastercycler PCR Thermocycler (10-minute initial denaturation at 94 °C, 30 cycles at 94 °C for 30 s, 55 °C for 60 s and 72 °C for 60 s, and final extension at 72 °C for 10 min). Amplified DNA was then run in 2% agarose gels, and only products of right size were selected for purification. DNA was directly purified from bacterial colonies using the QIAprep Miniprep kit (Qiagen) and sent to the C.A.I. laboratory to be run in a ABI Prism 3700 Genetic Analyzer (Applied Biosystems, Life Technologies). See Table S7 (Supporting information) for cloning alignments.

Consensus haplotype identification. Sequencing electropherograms were read using the program Mutation Surveyor (Demo) version 3.24 (SoftGenetics, LLC). Sequences were then aligned and consensus haplotypes were established using the following criteria:

Only those haplotypes that were repeated between different samples from the same individual using independent DNA extractions and amplifications were considered as endogenous.

Consensus haplotypes had to be congruent within the two fragments and with SNP typing.

Haplotypes matching those from laboratory staff, archaeologists or anthropologists were not considered.

Miscoding lesions were identified and removed from consensus haplotypes.

Jumping PCRs and cross-contaminations were individually estimated for each sample.

For further details about consensus haplotypes, see clones alignments (Table S7, Supporting information).

Haplogroup prediction and detection. Haplogroup prediction was based on mutations of the HVR-I according to Richards *et al.* (2000) and van Oven & Kayser (2009). Ambiguities were resolved by comparison of consensus haplotypes with available modern mtDNA sequences collected in public databases (mtDNA manager, Lee *et al.* 2008) and publications (Richards *et al.* 2000). Haplogroups were then confirmed by the amplification of SNPs in the mtDNA coding region using specific primers (Table S6, Supporting information), following the constantly upgraded phylogeny available developed by van Oven & Kayser (2009). PCR amplification, gel electrophoresis, purification and sequencing were performed as described earlier.

Ancient DNA data analysis

Summary statistics (F_{ST}). Following the recent studies of Haak *et al.* (2005) and Bramanti *et al.* (2009), we estimated pairwise genetic differentiation between three 'populations' using the F_{ST} parameter (Reynolds *et al.* 1983; Slatkin 1995) considering two overlapping fragments spanning 255 bp on mitochondrial HVR-I (16126–16379) and excluding primer annealing region, using Arlequin software, version 3.5.1.2. The three population groups were as follows: (i) EN (present study, $N = 13$, Can Sadurní, Chaves and Sant Pau del Camp sites), (ii) MN (Sampietro *et al.* 2007; $N = 11$, Camí de Can Grau site) and (iii) a modern samples from the same region (García *et al.* 2011; $N = 363$). The modern data set is a pool of 'Aragón' ($N = 119$), 'Catalonia–Aragón' ($N = 164$) and 'Catalonia' ($N = 80$) populations, fully described in the Supplementary Information from García *et al.* (2011).

Simulations and demographic models. Simulations with aDNA data were performed using the BayeSSC program (Excoffier *et al.* 2000) available at <http://www.stanford.edu/group/hadlylab/ssc/index.html>.

This program uses the Bayesian version of the serial coalescent algorithm to run simulations on genealogies backward in time according to the demographic model tested. We simulated three different data sets (Early Neolithic, Middle Neolithic, Modern Catalonia and Aragón) to match the observed data set in terms of sample sizes and time, as described earlier. In this framework, calibrated years BCE have been transformed to BP by adding 1950.

The previous studies of Haak *et al.* (2005) and Bramanti *et al.* (2009) assumed one model of panmixia across wide geographical areas and for tens of thousands of years. Here, we explored three demographic models to test genetic continuity in the studied area: TPM (total panmixia model), SM (split model) and

SDGM (split with differential growth model) (Fig. 2). For all three models, we followed the general framework developed in the previous studies, where an Upper Palaeolithic population was randomly sampled from a hypothetical African source population with a constant female effective size of 5000, representing the first modern humans that settled in Europe 'out of Africa' around 45 000 BP. The first model, named total panmixia model (TPM), was the simplest scenario and represented a random mating population settled in northeastern Spain, showing genetic continuity from the Upper Palaeolithic until the present. Even though this model is not realistic, it was tested to compare this unstructured model with the other two structured scenarios implemented here. TPM has been already used by Haak *et al.* (2005) and Bramanti *et al.* (2009) to test genetic continuity in Central Europe. R. Rasteiro and L. Chikhi (submitted) have shown that it was not consistent with Central European data and that structured models explained the data better. Here, we follow R. Rasteiro and L. Chikhi (submitted) and use the same demographic models. The second model tested (split model, SM) was designed to represent a simple departure from the TPM and allow for some structure. This scenario assumed that the original populations that colonized Europe 45 000 years ago split in two populations of the same size (demes 1 and 2) that met 7500 years ago, when the Neolithic spread into this region. The two branches can be seen as representing the Palaeolithic background of northeastern Spain and the Palaeolithic population in which the Neolithic arose in Near East. The third model (split with differential growth model, SDGM) was designed to reflect a different scenario with a larger Neolithic contribution. The SDGM was based on the same assumptions as the previous one, but one of the two demes represents the Near Eastern Neolithic population that starts growing around 10 000 BP at a higher rate than the other 'Palaeolithic' deme. As in the SM, the two demes join around 7500 BP, but the Neolithic branch (deme 2) is assumed to reach a population size 20 times larger than that of the other branch (deme 1). To summarize, the SDGM and SM differ in the effective

sizes of the two demes between 10 000 and 7500 BP: in the SM, demes 1 and 2 have the same effective size, while in the SDGM, the size of deme 1 is constrained to be 1/20th of deme 2 at time $t = 7500$ BP (Fig. 2).

For all these models, we assumed a sequence length of 255 bp, a fixed mutation rate of 7.5×10^{-6} per base pair and per generation (corresponding to 3×10^{-7} substitution per base pair and per year, Endicott & Ho 2008), a transition over transversion bias of 0.9841, a uniform gamma distribution of mutations with a shape parameter of 0.205 (Bandelt *et al.* 2006) and 25 years per generation. The modern population effective size (N_e) was set to two million. This number corresponds to around one-sixth of the census size of the region sampled and represents the effective population size of mitochondrial DNA. To estimate this value, we considered that (i) the population census size of the studied area is around 12 million (Statistics National Institute, Spain, <http://www.ine.es>), (ii) mtDNA represents only the female population, being around one-half of the total, and (iii) just around one-third of the female population reproduces (Cela-Conde & Ayala 2007). Effective population sizes at Upper Palaeolithic (UP; 45 000 BP) and Neolithic (N; 7500 BP) periods were estimated to have been around 700 for the former and 15 000 for the latter. These values were calculated from estimated population densities, around 0.064 individuals per km² for hunter-gatherers and 20 times this value for farmers (Steele *et al.* 1998; Alroy 2001). Given the uncertainty on these values, we followed Bramanti *et al.* (2009) and used wide uniform prior distributions, spanning from 10 to 5000 for UP and from 200 to 100 000 for N, by simulating 1 million genealogies per model. Growth rates were computed on the basis of UP and N effective population sizes sampled from the priors.

We then compared the simulated and observed F_{ST} ($simF_{ST}$ and $obsF_{ST}$, respectively) under an approximate Bayesian computation (ABC) approach (Beaumont *et al.* 2002; Blum & François 2009), using the *abc* package (Csilléry *et al.* 2010) implemented in the R 2.10.1 version (R Development Core Team, 2009). Parameters were estimated by retaining 10 000 simulated values, associ-

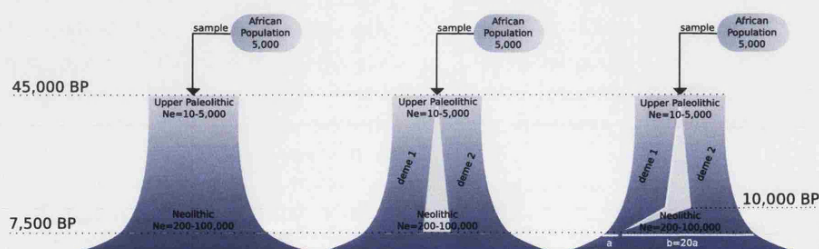


Fig. 2 Demographic models explored. From left to right: TPM, total panmixia model; split model, SM; split with differential growth model, SDGM; N_e , effective population sizes ranges.

ated with the shortest Euclidean distances (tolerance = 0.01, corresponding to the 1% closest data sets). Post rejection adjustments were made using the *neural-net* method and *logit* transformation (Blum & François 2009) available in the *abc* package (Table S4, Supporting information). On the basis of N_e posterior distributions obtained with the ABC framework, we set 10 values equally distributed within the UP size range (10–5000) and 30 values for the N size (20 ranging between 200 and 20 000 and 10 between 20 000 and 100 000). Then, we again used BayeSSC to run 1000 simulations for each of these size combinations ($10 \times 30 = 300$ combinations) and for each model (i.e. 900 000 simulations). Afterwards, we calculated for each parameter combination the probabilities of obtaining larger $\text{sim}F_{ST}$ values than observed ($\text{obs}F_{ST}$), under the three scenarios and plotted them using the *filled.countour* function in R version 2.10.1 (R Development Core Team, 2009), following previous studies using this approach (Bramanti *et al.* 2009; Malmström *et al.* 2009). Comparisons between $\text{sim}F_{ST}$ and $\text{obs}F_{ST}$ values within all $N-N_e$ values explored are shown in Fig. S1 (Supporting information). The ABC framework was also used to identify the model that best explains the observed F_{ST} , by computing posterior probabilities for each model (Beaumont 2008). To do this, we used the *postpr* function included in the *abc* package (Csilléry *et al.* 2010) in R version 2.10.1 (R Development Core Team, 2009), using a nonlinear conditional heteroscedastic method that applies a neural network approach (Blum & François 2009). This procedure was further validated by applying this approach to data sets for which the original model was known. We used 500 replicates within the 1 million simulated data set under each model (i.e. using the F_{ST} of the replicates as pseudo-observed F_{ST} values) and calculated how often the ABC procedure identified the correct model by a higher posterior probability (Table S4, Supporting information). We tested this approach when the three models were used together (i.e. choose one of three models) and in pairwise comparisons (TPM vs. SM and TPM vs. SDGM). Finally, we performed simulations by slightly changing the SM and SDGM models so as to have half of the aDNA samples in deme 1 or 2 by reducing the date at which the two demes join from 7500 to 7395 BP. Our aim was to determine whether the posterior probabilities of the structured models would change if we had had access to older aDNA samples.

Results

Haplotypes and haplogroups

We obtained consistent mitochondrial DNA results for 21 of 49 samples (42% success), corresponding to 13 of

22 individuals. The remaining 26 samples were discarded because of absence of results, lack of reproducibility or staff contamination. Samples recovered from caves (Can Sadurní and Chaves, 42% and 100% success, respectively) yielded a higher success rate than those from an open-air settlement (San Pau del Camp, 29% success), which agrees with earlier suggestions that caves offer stable temperature conditions all year round (de Torres *et al.* 2002) and protect samples from adverse climate conditions, such as rainfall (Hedges & Millard 1995).

Five different haplotypes were identified in the Can Sadurní archaeological site (Table 1). This information, combined with odontological age estimation (Table S1, Supporting information), allowed us to establish a minimum number of seven individuals, considering that samples with the same estimated age and haplotype were from the same individual. Altogether nine different haplotypes were found for the whole aDNA data set (Table 1). Most of them are currently widely distributed throughout Europe and belonged to the major European haplogroups H, K and U5. The first two haplogroups are currently distributed with high (H, ~46%) and moderate (K, ~6%) frequencies in Europe (Richards *et al.* 2000). U5 is currently present in around 9% of modern Europeans (Richards *et al.* 2000), and a recent aDNA analysis from Central European hunter-gatherers has shown that this haplogroup was present at a high frequency (64%) (Bramanti *et al.* 2009).

Three of the haplotypes were represented in more than one individual: 16224C, 16311C (one individual at each site), 16223T, 16362C (two individuals from the same site) and 16147T, 16223T, 16362C (two individuals from two different sites). These last two were assigned through SNP typing to haplogroup N*. This haplogroup was absent in MN and in the modern sample from northeastern Iberia (Table S2, Supporting information), but it has been previously detected in modern populations from the Near East and Eastern Europe (Table S3, Supporting information) and in a 25 000-year-old Cro-Magnon specimen from the Paglicci cave, southern Italy (Caramelli *et al.* 2003). However, haplogroup N* appears to be rare in modern populations as it could not be found in a combined search in most common databases (Richards *et al.* 2000; Lee *et al.* 2008; van Oven & Kayser 2009). We also identified the rare X1 branch in one Eastern Iberian Neolithic sample. Haplogroup X is found at low frequencies (2–3%) in Europeans, Near Easterners, North Africans and native Americans (Reidla *et al.* 2003). Subclade X1 has an early coalescence time ($42\,900 \pm 11\,900$ BP) and is currently restricted to Northern and Eastern Africa and the Near East (Table S3, Supporting information) (Reidla *et al.*

Table 1 Mitochondrial DNA haplotypes and haplogroups of the studied samples

Site	Period	Date	Specimen	Haplotype	Single nucleotide polymorphisms typing	Haplogroup
Can Sadurní	Early Neolithic (Cardial)	5475–5305 cal. BCE	CSA0511	16223T, 16362C	5178C, 4833A, 10873T, 10398A, 10400C, 10238T	N*
Can Sadurní	Early Neolithic (Cardial)	5475–5305 cal. BCE	CSA09	16223T, 16362C	5178C, 4833A, 10873T, 10398A, 10400C, 10238T	N*
Can Sadurní	Early Neolithic (Cardial)	5475–5305 cal. BCE	CSA152223	16224C, 16311C	10550G	K
Can Sadurní	Early Neolithic (Cardial)	5475–5305 cal. BCE	CSA16	16362C	7028C	H
Can Sadurní	Early Neolithic (Cardial)	5475–5305 cal. BCE	CSA24	16136C, 16192T, 16270T	14766T, 3197C	U5
Can Sadurní	Early Neolithic (Cardial)	5475–5305 cal. BCE	CSA26	16189C, 16189C (het), 16223T, 16278T	7028T, 1719C, 6371T	X1
Can Sadurní	Early Neolithic (Cardial)	5475–5305 cal. BCE	CSA29	16147T, 16223T, 16362C	10398A, 10400C, 10873T, 10238T	N*
Chaves	Early Neolithic (Cardial)	5329–4999 cal. BCE	1CH0102	16224C, 16311C	10550G	K
Chaves	Early Neolithic (Cardial)	5329–4999 cal. BCE	2CH0102	CRS	7028C	H
Chaves	Early Neolithic (Cardial)	5329–4999 cal. BCE	3CH01	16129A	7028C	H
Sant Pau del Camp	Late Early Neolithic	4250–3700 cal. BCE	6SP0102	16224C, 16311C	10550G	K
Sant Pau del Camp	Late Early Neolithic	4250–3700 cal. BCE	26SP0102	16218T, 16328A, (16362C)	7028C	H20
Sant Pau del Camp	Late Early Neolithic	4250–3700 cal. BCE	27SP0102	16147T, 16223T, 16362C	10873T, 7028T, 10238T	N*

In this table, mtDNA results are reported (haplotype, haplogroup and coding regions amplified), in correspondence of each specimen name, archaeological site and age.

2003). Only the branch X2 has been previously detected in ancient samples from French MN and Central European Corded Ware Culture (Haak *et al.* 2008; Deguilloux *et al.* 2011).

Genetic distances and simulation results

Our EN samples were compared with published data from MN (Sampietro *et al.* 2007) and with modern populations (García *et al.* 2011) from the same geographical area. Differences in haplotype composition between EN and modern samples were reflected in the high and highly significant pairwise F_{ST} value ($F_{ST} = 0.131$, $P < 10^{-5}$). A similarly high and significant value was observed between EN and MN populations ($F_{ST} = 0.101$, $P < 10^{-5}$). However, the MN and modern Iberian samples presented a lower and nonsignificant pairwise F_{ST} value ($F_{ST} = 0.032$, $P = 0.072$), suggesting significant drift between the Early and Middle Neolithic but less between the latter and the present. Comparisons between $simF_{ST}$ and $obsF_{ST}$ values are shown in Fig. 3. The left-hand panels (a.1–3) represent the comparison between MN and modern populations, whereas the central (b.1–3) and right-hand (c.1–3) panels correspond to the EN vs. modern-day and EN vs. MN samples. The top panels (a–c.1) correspond to the results under the TPM, whereas the central (a–c.2) and bottom (a–c.3) panels correspond to the SM and SDGM, respectively. The three models can explain the $obsF_{ST}$ between MN and modern-day samples for all population sizes tested (panels a.1–3 and Fig. S1, Supporting information). However, the $obsF_{ST}$ value between EN and modern populations can be only explained under small Neolithic population sizes (<1000 under the TPM and up to 2000 for SM and SDGM, panels b.1–3). When comparing EN and MN populations (panels c.1–3), it can be seen that structured models explain the observed statistics for a wider range of combinations of Neolithic effective population sizes than the unstructured model (TPM). Thus, to explain the three $obsF_{ST}$ values, the TPM only allows a very limited ranges of parameter values.

Table 2 shows the posterior probabilities for each of the three demographic models, estimated under an ABC model choice framework. The lowest posterior probability (16%) value was obtained for TPM, whereas posterior probability values were higher (>40%) for the structured models (SM and SDGM) when we compared the three models at once. When we compare the models by pairs, the posterior probabilities of the structured models are always much higher (>72%) than that of the TPM (<28%), whereas the difference between the two structured models is limited. These results clearly favour the structured models over the TPM. It is

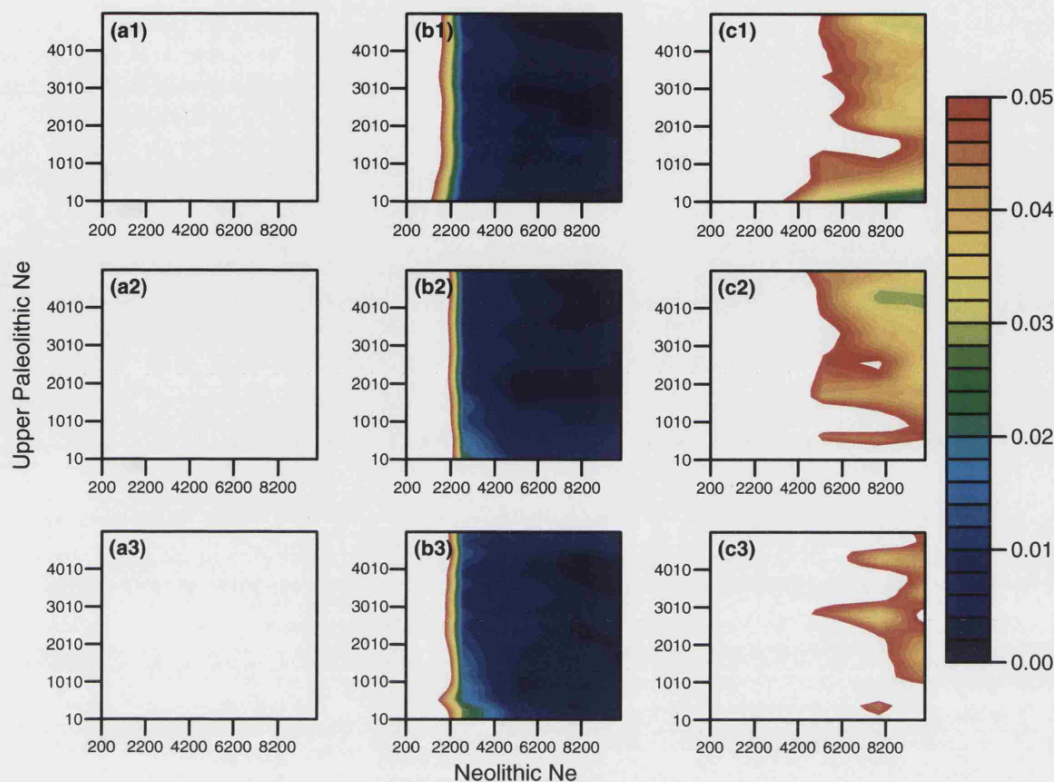


Fig. 3 Probabilities of obtaining simulated F_{ST} greater than observed. (a–c.1) TPM model, (a–c.2) SM model, (a–c.3) SDGM model, (a.1–3) MN and modern-day populations F_{ST} ; (b.1–3) EN and modern-day populations F_{ST} ; (c.1–3) MN and EN populations F_{ST} . Values higher than 0.05 are in white. $N-N_e$ range represented is up to 10 000. Plots with all $N-N_e$ values explored are available in Fig. S1 (Supporting information). UP- N_e , Upper Palaeolithic effective population size; $N-N_e$, Neolithic effective population size; TPM, total panmixia model; split model, SM; split with differential growth model, SDGM; MN, Middle Neolithic; EN, Early Neolithic.

Table 2 Posterior probabilities of the three models (TPM, SM and SDGM)

Models	Posterior model probabilities (%)
Total panmixia (TPM)	16.1
Split (SM)	40.4
Split with differential growth (SDGM)	43.4

Posterior probabilities of the three models estimated using the approximate Bayesian computation (ABC) scheme described in the text. Posterior probabilities were obtained using the *neurafnet* method with a tolerance = 0.1, as implemented in the R *abc* package.

interesting to note that the support for the TPM is much lower than obtained in the validation under any of the models (Table S4, Supporting information), which again suggests that the TPM is a very unlikely model. Finally, when we modified the SM and SDGM to have aDNA samples in deme 1 or 2, we found that the posterior distributions for the structured models increased significantly (>93%).

Discussion

For several decades, the Neolithic transition has been at the centre of ongoing controversies among archaeologists and geneticists, particularly since the advent of mtDNA and Y chromosome data in the 1990s (Richards *et al.* 1996; Barbujani *et al.* 1998; Semino *et al.* 2000; Chikhi *et al.* 2002; Richards 2003). Beyond the methodological disagreements among authors, it is important to note that several genetic studies, while defending one model of Neolithic spread, have acknowledged the possibility that different regions may have witnessed different processes (Barbujani & Chikhi 2006). They have thus called for regional studies to determine whether it is possible to separate demic from cultural processes using both modern and aDNA. Our study addresses exactly these issues by presenting the first aDNA evidence from the earliest Neolithic communities reaching the Iberian Peninsula.

This study demonstrates that it is possible to recover ancient endogenous DNA from temperate environments such as northeastern Iberia. Moreover, the results obtained suggest that aDNA preservation depends

more on depositional conditions than on sample age, in agreement with previous observations (Nielsen-Marsh & Hedges 2000).

Haplotype and haplogroup composition of EN and MN populations allow us to infer possible genetic relationships between populations and/or archaeological sites. For example, shared haplotypes among Can Sadurní and Sant Pau del Camp could point at a certain degree of genetic continuity during the Neolithic in northeastern Iberia. Regarding the haplogroup composition, modern and MN populations differ from EN samples, mostly due to the presence of rare haplogroups N* and X1 in the latter. These differences are reflected in high $obsF_{ST}$. Similarly, high $simF_{ST}$ values could also be observed for a wide range of parameter values in the two structured models (SM and SDGM) but very rarely in TPM (Fig. 3). Within this range, the effect of genetic drift could have produced the loss of rare haplogroups N* and X1, as has been suggested for haplogroup N1a in Central Europe (Barbujani & Bertorelle 2001; Barbujani & Chikhi 2006; Haak *et al.* 2010; R. Rasteiro and L. Chikhi, submitted). This explains why our simulations favour the models considering previous population structure over an unstructured model (TPM) (Table 2). While the latter model is clearly too simplistic, it has been the one used in recent aDNA studies (Haak *et al.* 2005; Bramanti *et al.* 2009). Geographical and chronological origins of N* and X1 rare lineages found in Iberian EN are difficult to trace with current information. These are, however, currently present in the Near East (Table S3, Supporting information). The possibility that these haplogroups were carried by Neolithic immigrants from the Near East seems to be supported from an archaeological point of view. For example, affinities in certain burial rituals (Hodder 2007; Utrilla *et al.* 2008) have been detected along the Neolithic spread in Europe, including the Cardial culture. In this particular case, individual 1CH0102 (Chaves site) showed a Near Eastern burial ritual. Other haplogroups found at high frequency in our sample, such as H and K, could have also been introduced together with the Neolithic expansion (Barbujani *et al.* 1998; Barbujani & Chikhi 2006). Further aDNA data from Near Eastern Neolithic and Iberian Mesolithic specimens are needed, as well as a wider analysis of specimens belonging to the Cardial culture (western Italy, southern France, Mediterranean and Atlantic coast of the Iberian Peninsula) to better identify haplotypes and haplogroups that were present in different locations spanning the Near East and all of Europe. Our results allowed us to reject the TPM but did not allow us to easily separate the two structured models. We expect that the incorporation of new aDNA data from older northeastern Iberian Mesolithic and Near Eastern Neolithic sites could help distinguish

them. Most of the samples used here were relatively recent in relation to the date at which demes 1 and 2 joined. However, when we modified this date to simulate the availability of more ancient aDNA, we found that it improved our ability to separate the models.

To conclude, we can tentatively propose a scenario that could explain our results, namely (i) the presence of currently rare haplogroups (N* and X1) in EN samples, (ii) high genetic drift during the period between the Early and Middle Neolithic, (iii) genetic affinities between EN and the Near East area and (iv) cultural connections with the Near East. This scenario would require that genetic drift played an important role at the beginning of the Neolithic with Near Eastern connections, hence pointing at a succession of pioneer colonization events from the Near East, which might point at other migration events along the Mediterranean, which might be identified in future studies.

Acknowledgements

This work was supported by CGL2006-07828/BOS and CGL2009-07959 research projects (Ministry of Science and Innovation (MICINN), Spanish Government). Human resources were funded by an FPU grant (ref. AP2006-01586, Spanish Government) for C.G., a grant (ref. SFRH/BD/30821/2006, Fundação para a Ciência e Tecnologia, Portugal) for R.R. and a MICINN researcher contract 'Juan de la Cierva' for E.F. (partially supported by the European Social Fund). Replications analyses in Bordeaux have been partly supported by the CNRS (Centre National de Recherches Scientifiques). Funding was provided to LC by the 'Laboratoire d'Excellence (LABEX)' entitled TULIP (ANR -10-LABX-41). We would like to thank all archaeologists and anthropologists who collaborated in this work: Anna Blasco, María Josefa Villalba, María Saña, Teresa Cabellos, Jordi Ruiz (Can Sadurní); Vicente Baldellou (Chaves); Josep Anfruns, Alejandro Pérez-Pérez, Mohammad Alrousan, Ferrán Estebaranz, Laura Martínez (Sant Pau del Camp). Thanks to Juan Álvarez for developing the sequence analysis program used here. The simulations were carried out in the HPC resources from CALMIP, Toulouse, France (Grant 2010-P1038).

References

- Alroy J (2001) A multispecies overkill simulation of the end-Pleistocene megafaunal mass extinction. *Science*, **292**, 1893–1896.
- Bandelt H-J, Macaulay V and Richards M (eds) (2006) Estimation of mutation rates and coalescence times: some caveats. In: *Human Mitochondrial DNA and the Evolution of Homo sapiens*, vol. 18(part 1), pp. 47–90. Springer, Berlin, Heidelberg.
- Barbujani G, Bertorelle G (2001) Genetics and the population history of Europe. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 22–25.
- Barbujani G, Chikhi L (2006) DNAs from the European Neolithic. *Heredity*, **97**, 84–85.

- Barbujani G, Bertorelle G, Chikhi L (1998) Evidence for Palaeolithic and Neolithic gene flow in Europe. *American Journal of Human Genetics*, **62**, 488–492.
- Beaumont MA (2008) Joint determination of topology, divergence time, and immigration in population trees. In: *Simulation, Genetics, and Human Prehistory* (eds Matsumura S, Forster P and Renfrew C), pp. 135–154. McDonald Institute for Archaeological Research, Cambridge.
- Beaumont MA, Zhang W, Balding DJ (2002) Approximate Bayesian computation in population genetics. *Genetics*, **162**, 2025–2035.
- Beaumont MA, Nielsen R, Robert C *et al.* (2010) In defence of model-based inference in phylogeography. *Molecular Ecology*, **19**, 436–446.
- Bernabeu J (1997) Indigenism and migrationism. The neolithisation of the Iberian peninsula. *Documenta Praehistorica*, **XXIV**, 1–17.
- Blasco A, Edo M, Villalba MJ, Saña M (2005) Primeros datos sobre la utilización sepulcral de la Cueva de Can Sadurní (Begues, Baix Llobregat) en el Neolítico Cardial. In: *Actas del III Congreso del Neolítico en la Península Ibérica* (eds Arias P, Otañón R and García-Moncó C), pp. 625–634. Servicio de Publicaciones, Universidad de Cantabria, Santander.
- Blum MGB, François O (2009) Non-linear regression models for approximate Bayesian computation. *Statistics and Computing*, **20**, 63–73.
- Bocquet-Appel J-P, Bar-Yosef O (2008) *The Neolithic Demographic Transition and its Consequences*. Springer, Dordrecht, The Netherlands.
- Bramanti B, Thomas MG, Haak W *et al.* (2009) Genetic discontinuity between local hunter-gatherers and Central Europe's first farmers. *Science*, **326**, 137–140.
- Brotherton P, Endicott P, Sanchez JJ *et al.* (2007) Novel high-resolution characterization of ancient DNA reveals C > U-type base modification events as the sole cause of post mortem miscoding lesions. *Nucleic Acids Research*, **35**, 5717–5728.
- Caramelli D, Lalueza-Fox C, Vernesi C *et al.* (2003) Evidence for a genetic discontinuity between Neandertals and 24,000-year-old anatomically modern Europeans. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 6593–6597.
- Cela-Conde CJ, Ayala FJ (2007) *Human Evolution: Trails from the Past*, 1st edn. Oxford University Press, New York.
- Chikhi L, Nichols RA, Barbujani G, Beaumont MA (2002) Y genetic data support the Neolithic demic diffusion model. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 11008–11013.
- Childe VG (1964) *The Dawn of European Civilization*, 6th edn. Vintage Books, New York.
- Csilléry K, Blum MGB, Gaggiotti OE, François O (2010) Approximate Bayesian computation (ABC) in practice. *Trends in Ecology & Evolution*, **25**, 410–418.
- Deguiloux M-F, Soler L, Pemonge M-H, Scarre C, Joussaume R, Laporte L (2011) News from the west: ancient DNA from a French megalithic burial chamber. *American Journal of Physical Anthropology*, **144**, 108–118.
- Dupanloup I, Bertorelle G, Chikhi L, Barbujani G (2004) Estimating the impact of prehistoric admixture on the genome of Europeans. *Molecular Biology and Evolution*, **21**, 1361–1372.
- Endicott P, Ho SYW (2008) A Bayesian evaluation of human mitochondrial substitution rates. *American Journal of Human Genetics*, **82**, 895–902.
- Excoffier L, Novembre J, Schneider S (2000) Computer note. SIMCOAL: a general coalescent program for the simulation of molecular data in interconnected populations with arbitrary demography. *Journal of Heredity*, **91**, 506–509.
- Fernández J, Gómez M (2009) Climate change and population dynamics during the Late Mesolithic and the Neolithic transition in Iberia. *Documenta Praehistorica*, **XXVIII**, 67–96.
- García O, Fregel R, Larruga JM *et al.* (2011) Using mitochondrial DNA to test the hypothesis of a European post-glacial human recolonization from the Franco-Cantabrian refuge. *Heredity*, **106**, 37–45.
- Gilbert MTP, Binladen J, Miller W *et al.* (2007) Recharacterization of ancient DNA miscoding lesions: insights in the era of sequencing-by-synthesis. *Nucleic Acids Research*, **35**, 1–10.
- Haak W, Forster P, Bramanti B *et al.* (2005) Ancient DNA from the first European farmers in 7500-year-old Neolithic sites. *Science*, **310**, 1016–1018.
- Haak W, Brandt G, de Jong HN *et al.* (2008) Ancient DNA, Strontium isotopes, and osteological analyses shed light on social and kinship organization of the Later Stone Age. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 18226–18231.
- Haak W, Balanovsky O, Sanchez JJ *et al.* (2010) Ancient DNA from European early Neolithic farmers reveals their near eastern affinities. *PLoS Biology*, **8**, e1000536.
- Hedges REM, Millard AR (1995) Bones and groundwater: towards the modelling of diagenetic processes. *Journal of Archaeological Science*, **22**, 155–164.
- Hodder I (2007) *Excavating Catalhoyuk: South, North and KOPAL Area Reports from the 1995–99 Seasons*. McDonald Institute for Archaeological Research, Cambridge.
- Lee HY, Song I, Ha E, Cho S-B, Yang WI, Shin K-J (2008) mtDNAMAN: a web-based tool for the management and quality analysis of mitochondrial DNA control-region sequences. *BMC Bioinformatics*, **9**, 483.
- Malmström H, Gilbert MTP, Thomas MG *et al.* (2009) Ancient DNA reveals lack of continuity between Neolithic hunter-gatherers and contemporary Scandinavians. *Current Biology*, **19**, 1758–1762.
- Molist M, Vicente O, Farré R (2008) El jaciment de la caserna de Sant Pau del Camp: aproximació a la caracterització d'un assentament del neolític antic. *Quarhis*, **4**, 14–24.
- Montes L, Utrilla P (2008) Le Paléolithique Supérieur dans la moyenne vallée de l'Ebre. *L'Anthropologie*, **112**, 168–181.
- Nielsen-Marsh CM, Hedges REM (2000) Patterns of diagenesis in bone I: the effects of site environments. *Journal of Archaeological Science*, **27**, 1139–1150.
- van Oven M, Kayser M (2009) Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Human Mutation*, **30**, E386–E394.
- Price TD (2000) *Europe's First Farmers*. Cambridge University Press, New York.
- R Development Core Team (2009) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at: <http://www.R-project.org>.

- Reidla M, Kivisild T, Metspalu E *et al.* (2003) Origin and diffusion of mtDNA haplogroup X. *The American Journal of Human Genetics*, **73**, 1178–1190.
- Reynolds J, Weir BS, Cockerham CC (1983) Estimation of the coancestry coefficient: basis for a short-term genetic distance. *Genetics*, **105**, 767–779.
- Richards M (2003) The Neolithic invasion of Europe. *Annual Review of Anthropology*, **32**, 135–162.
- Richards M, Côrte-Real H, Forster P *et al.* (1996) Paleolithic and neolithic lineages in the European mitochondrial gene pool. *American Journal of Human Genetics*, **59**, 185–203.
- Richards M, Macaulay V, Hickey E *et al.* (2000) Tracing European founder lineages in the near Eastern mtDNA pool. *American Journal of Human Genetics*, **67**, 1251–1276.
- Rohland N, Hofreiter M (2007) Ancient DNA extraction from bones and teeth. *Nature Protocols*, **2**, 1756–1762.
- Sampietro ML, Lao O, Caramelli D *et al.* (2007) Palaeogenetic evidence supports a dual model of Neolithic spreading into Europe. *Proceedings of the Royal Society B: Biological Sciences*, **274**, 2161–2167.
- Semino O, Passarino G, Oefner PJ *et al.* (2000) The genetic legacy of paleolithic *Homo sapiens sapiens* in Extant Europeans: a Y chromosome perspective. *Science*, **290**, 1155–1159.
- Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, **139**, 457–462.
- Soares P, Achilli A, Semino O *et al.* (2010) The archaeogenetics of Europe. *Current Biology*, **20**, R174–R183.
- Steele J, Adams J, Sluckin T (1998) Modelling Paleoindian dispersals. *World Archaeology*, **30**, 286–305.
- de Torres T, Ortiz JE, Llamas FJ, Canoira L, Julia R, García-Martínez MJ (2002) Bear dentine aspartic acid racemization analysis: a proxy for the dating of Pleistocene cave infills. *Archaeometry*, **44**, 417–426.
- Utrilla P, Lorenzo JL, Baldellou V, Sopena MC, Ayuso P (2008) Enterramiento masculino en fosa, cubierto de cantos rodados, en el neolítico antiguo de la cueva de Chaves. In: *IV Congreso del Neolítico Peninsular* (eds Hernández MS, Soler JA and López JA), vol. 2, pp. 131–140. Museo Arqueológico de Alicante (MARQ), Alicante.
- Zilhão J (2001) Radiocarbon evidence for maritime pioneer colonization at the origins of farming in west Mediterranean Europe. *Proceedings of the National Academy of Sciences*, **98**, 14180–14185.
- Zvelebil M (2001) The agricultural transition and the origins of Neolithic society in Europe. *Documenta Praehistorica*, **XXVIII**, 1–29.

CG, EF, MT, MFD, MHP and EAP are interested in using ancient DNA genetic analyses to gain insights into archaeological and demographic aspects of ancient human populations, particularly in the genetic impact of the Neolithic. CG, RR and LC are interested in applying model-based approaches to study the demographic past of populations, using either ancient and/or modern genetic data. PU, ME and MM are interested in the archaeological and demographic Prehistory of European, Near Eastern and Iberian populations.

Data accessibility

Sequence alignments are available in the Supporting information.

Supporting information

Additional supporting information may be found in the online version of this article.

Fig. S1 Plots with $N-N_e$ values explored.

Table S1 Samples description.

Table S2 Haplogroup frequencies in studied populations.

Table S3 Haplogroup N* and X1 frequencies in the Near East.

Table S4 Posterior distribution of population sizes and power to recover the true model.

Table S5 Staff's mtDNA.

Table S6 Primer sequences.

Table S7 Cloned sequences alignments.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

SUPPLEMENTARY MATERIAL

TABLE S1

Sample description

Site	Province	Specimen	Sample	Period	Date	Date reference	Age	Sample type
Can Sadurní	Barcelona	CSA0511	CSA05	Early Neolithic (Cardial)	5,475-5,305 cal. BC	Blasco et al. 2005	adult	canine tooth
Can Sadurní	Barcelona	CSA0511	CSA11	Early Neolithic (Cardial)	5,475-5,305 cal. BC	Blasco et al. 2006	adult	incisor tooth
Can Sadurní	Barcelona	CSA09	CSA09	Early Neolithic (Cardial)	5,475-5,305 cal. BC	Blasco et al. 2007	infant	molar tooth
Can Sadurní	Barcelona	CSA152223	CSA15	Early Neolithic (Cardial)	5,475-5,305 cal. BC	Blasco et al. 2008	adult/sub-adult	molar tooth
Can Sadurní	Barcelona	CSA152223	CSA22	Early Neolithic (Cardial)	5,475-5,305 cal. BC	Blasco et al. 2009	adult/sub-adult	premolar tooth
Can Sadurní	Barcelona	CSA152223	CSA23	Early Neolithic (Cardial)	5,475-5,305 cal. BC	Blasco et al. 2010	adult/sub-adult	premolar tooth
Can Sadurní	Barcelona	CSA16	CSA16	Early Neolithic (Cardial)	5,475-5,305 cal. BC	Blasco et al. 2011	adult/sub-adult	incisor tooth
Can Sadurní	Barcelona	CSA24	CSA24	Early Neolithic (Cardial)	5,475-5,305 cal. BC	Blasco et al. 2012	infant	molar tooth
Can Sadurní	Barcelona	CSA26	CSA26	Early Neolithic (Cardial)	5,475-5,305 cal. BC	Blasco et al. 2013	infant	molar tooth
Can Sadurní	Barcelona	CSA29	CSA29	Early Neolithic (Cardial)	5,475-5,305 cal. BC	Blasco et al. 2014	infant	canine tooth
Chaves	Huesca	1CH0102	1CH01	Early Neolithic (Cardial)	5,329-4,999 cal BC	Utrilla et al. 1999	adult	cranial fragment
Chaves	Huesca	1CH0102	1CH02	Early Neolithic (Cardial)	5,329-4,999 cal BC	Utrilla et al. 2000	adult	rib fragment
Chaves	Huesca	2CH0102	2CH01	Early Neolithic (Cardial)	5,329-4,999 cal BC	Utrilla et al. 2001	adult	molar tooth
Chaves	Huesca	2CH0102	2CH02	Early Neolithic (Cardial)	5,329-4,999 cal BC	Utrilla et al. 2002	adult	molar tooth
Chaves	Huesca	3CH01	3CH01	Early Neolithic (Cardial)	5,329-4,999 cal BC	Utrilla et al. 2003	adult	molar tooth
Sant Pau del Camp	Barcelona	19SP0102	19SP01	Late Early Neolithic	4,250-3,700 cal BC	Molist et al. 2008	adult	molar tooth
Sant Pau del Camp	Barcelona	19SP0102	19SP02	Late Early Neolithic	4,250-3,700 cal BC	Molist et al. 2009	adult	premolar tooth
Sant Pau del Camp	Barcelona	26SP0102	26SP01	Late Early Neolithic	4,250-3,700 cal BC	Molist et al. 2010	adult	molar tooth
Sant Pau del Camp	Barcelona	26SP0102	26SP02	Late Early Neolithic	4,250-3,700 cal BC	Molist et al. 2011	adult	molar tooth
Sant Pau del Camp	Barcelona	27SP0102	27SP01	Late Early Neolithic	4,250-3,700 cal BC	Molist et al. 2012	infant	incisor tooth
Sant Pau del Camp	Barcelona	27SP0102	27SP02	Late Early Neolithic	4,250-3,700 cal BC	Molist et al. 2013	infant	molar tooth

References

Blasco, A., Edo, M., Villalba, M. J. & Saña, M. 2005 Primeros datos sobre la utilización sepulcral de la Cueva de Can Sadurní (Begues, Baix Llobregat) en el Neolítico Cardial. In *Actas del III Congreso del Neolítico en la Península Ibérica.*, Santander: Servicio de Publicaciones, Universidad de Cantabria.

Utrilla, M. del P. & Calvo, M. J. 1999 Cultura material y arte rupestre "levantino": La aportación de los yacimientos aragoneses a la cuestión cronológica. Una revisión del tema en el año 2000. *Bolskan* 16, 39-70.

Molist, M., Vicente, O. & Farré, R. 2008 El jaciment de la caserna de Sant Pau del Camp: aproximació a la caracterització d'un assentament del neolític antic. *Quarhis* 4, 14_24.

TABLE S2

Haplogroups frequencies in Early Neolithic samples (present study), Middle Neolithic Samples (Sampietro et al. 2007) and Modern population samples from the same area (García et al. 2011).

		Population					
		Modern (%)	Modern (N)	Middle Neolithic (%)	Middle Neolithic (N)	Early Neolithic (%)	Early Neolithic (N)
Haplogroup	H/HV/U*/R*	42,70	155	36,36	4	30,77	4
	N*	-	-	-	-	30,77	4
	K	6,06	22	-	-	23,08	3
	V/HV0	4,96	18	-	-	-	-
	U4	0,83	3	9,09	1	-	-
	U5	8,54	31	-	-	7,69	1
	U6	1,38	5	-	-	-	-
	T	10,19	37	18,18	2	-	-
	J	9,64	35	18,18	2	-	-
	N1a	0,28	1	-	-	-	-
	I	1,38	5	9,09	1	-	-
	W	2,48	9	9,09	1	-	-
	X	2,48	9	-	-	7,69	1
	others	9,09	33	-	-	-	-
	TOTAL	100,00	363	100,00	11	100,00	13

TABLE S3

Haplogroup N* frequencies in the Near East and Eastern Europe

Population name	N* (%)	Region	Population	N	References
Saudi Arabia	0,18	Saudi Arabia	Arab	553	Abu Amero et al. 2008
Oman	2,86	Oman	Arab	105	Rowold et al. 2007
United Arab Emirates	9,16	Union Emirates	Arab Arab	131	Rowold et al. 2007
Jordan	0,54	Jordan	Arab	184	González et al. 2008, Rowold et al. 2007
Ethiopia	1,48	Ethiopian	Ethiopian	270	Kivisild et al. 2004
Iran	1,58	Iran	Iranian	507	Richards et al. 2000, Metspalu et al. 2004, Comas et al. 2004, Quintana-Murci et al. 2004
Syria	0,85	Syria	Syrian	118	Vernesi et al. 2001, Richards et al. 2000
Palestinian	0,81	Israel	Palestinian	247	Shen et al. 2004, Richards et al. 2000, Behar et al. 2008, Di Rienzo and Wilson 1991
Druze	0,22	Near East	Druze	452	Macaulay et al. 1999, Shen et al. 2004, Slush et al. 2008, Behar et al. 2008
Turkmenistan	1,64	Turkmenistan	Turkmen	61	Quintana-Murci et al. 2004, Comas et al. 2004
Near Eastern Jews	1,68	Yemen, Iraq, Syria, Iran, Kurdistan, Uzbekistan, Afghanistan	Jew	476	Shen et al. 2004, Richards et al. 2000, Behar et al. 2008, Di Rienzo and Wilson 1991, Picornell et al. 2006
Serbia	1,71	Serbia	Serbian	117	Cvjetan et al. 2004
Askhenazi Jews	0,32	Austria, Hungary, Byelorussia, Czech Republic, Estonia, France, Germany, Latvia, Lithuania, Moldavia, Netherlands, Poland, Romania, Russia, Switzerland, Ukraine	Jew	623	Behar et al. 2004, Behar et al. 2006, Shen et al. 2004, Picornell et al. 2006

Haplogroup X1 frequencies in the Near East

Population name	X1 (%)	Region	Population	N	References
Near Eastern Jews	0,03	Iraq, Iran	Jewish	135	Behar et al. 2008
Caucasian Jews	0,02	Azerbaijan	Jewish	58	Behar et al. 2008
Palestinian	0,01	Israel	Palestinian	110	Behar et al. 2008
Askhenazi Jews	0,03	Austria, Hungary	Jewish	29	Behar et al. 2006
Ethiopia	0,01	Ethiopian	Ethiopian	270	Kivisild et al. 2004
Kuwait	0,50	Kuwait	Kwaitis	202	Reidla et al. 2003
Druze	15,60	Israel	Druze	45	Macaulay et al. 1999
Jordan	0,50	Jordan	Arab	202	Reidla et al. 2003
Iraq	0,90	Iraq	Iraqian	116	Richards et al. 2000
Iran	0,20	Iran	Iranian	440	Reidla et al. 2003

References

Abu-Amero KK, González AM, Larruga JM, Bosley TM, and Cabrera VM. Eurasian and African mitochondrial DNA influences in the Saudi Arabian population. *BMC Evol Biol* 7:32-32.

Behar DM, Metspalu E, Kivisild T, Achilli A, Hadid Y, Tzur S, Pereira L, Amorim A, Quintana-Murci L, Majamaa K, et al. 2006. The matrilineal ancestry of Ashkenazi Jewry: portrait of a recent founder event. *Am. J. Hum. Genet* 78:487-497.

Behar DM, Metspalu E, Kivisild T, Rosset S, Tzur S, Hadid Y, Yudkovsky G, Rosengarten D, Pereira L, Amorim A, et al. 2008. Counting the founders: the matrilineal genetic ancestry of the Jewish Diaspora. *PLoS ONE* 3:e2062.

Comas D, Plaza S, Wells RS, Yuldaseva N, Lao O, Calafell F, and Bertranpetit J. 2004. Admixture, migrations, and dispersals in Central Asia: evidence from maternal DNA lineages. *Eur. J. Hum. Genet* 12:495-504.

Cvjetan S, Tolk H-V, Lauc LB, Colak I, Dordević D, Efremovska L, Janićijević B, Kvesić A, Klarić IM, Metspalu E, et al. 2004. Frequencies of mtDNA haplogroups in southeastern Europe- Croatians, Bosnians and Herzegovinians, Serbians, Macedonians and Macedonian Romani. *Coll Antropol* 28:193-198.

Di Rienzo A, and Wilson AC. 1991. Branching pattern in the evolutionary tree for human mitochondrial DNA. *Proc. Natl. Acad. Sci. U.S.A* 88:1597-1601.

González AM, Karadsheh N, Maca-Meyer N, Flores C, Cabrera VM, and Larruga JM. 2008. Mitochondrial DNA variation in Jordanians and their genetic relationship to other Middle East populations. *Ann. Hum. Biol* 35:212-231.

Kivisild T, Reidla M, Metspalu E, Rosa A, Brehm A, Pennarun E, Parik J, Geberhiwot T, Usanga E, and Vilems R. 2004. Ethiopian mitochondrial DNA heritage: tracking gene flow across and around the gate of tears. *Am. J. Hum. Genet* 75:752-770.

Macaulay V, Richards M, Hickey E, Vega E, Cruciani F, Guida V, Scozzari R, Bonn -Tamir B, Sykes B, and Torroni A. 1999. The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. *Am J Hum Genet* 64:232-249.

Picornell A, Gim nez P, Castro JA, and Ramon MM. 2006. Mitochondrial DNA sequence variation in Jewish populations. *Int. J. Legal Med* 120:271-281.

Quintana-Murci L, Chaix R, Wells RS, Behar DM, Sayar H, Scozzari R, Rengo C, Al-Zahery N, Semino O, Santachiara-Benerecetti AS, et al. 2004. Where West Meets East: The Complex mtDNA Landscape of the Southwest and Central Asian Corridor. *Am J Hum Genet* 74:827-845.

Richards M, Macaulay V, Hickey E, Vega E, Sykes B, Guida V, Rengo C, Sellitto D, Cruciani F, Kivisild T, et al. 2000. Tracing European founder lineages in the Near Eastern mtDNA pool. *Am. J. Hum. Genet* 67:1251-1276.

Rowold DJ, Luis JR, Terreros MC, and Herrera RJ. 2007. Mitochondrial DNA gene flow indicates preferred usage of the Levant Corridor over the Horn of Africa passageway. *J. Hum. Genet* 52:436-447.

Shen P, Lavi T, Kivisild T, Chou V, Sengun D, Gefel D, Shpirer I, Woolf E, Hillel J, Feldman MW, et al. 2004. Reconstruction of patrilineages and matrilineages of Samaritans and other Israeli populations from Y-chromosome and mitochondrial DNA sequence variation. *Hum. Mutat* 24:248-260.

Shlush LI, Behar DM, Yudkovsky G, Templeton A, Hadid Y, Basis F, Hammer M, Itzkovitz S, and Skorecki K. 2008. The Druze: A Population Genetic Refugium of the Near East. *PLoS ONE* 3.

Vernesi C, Di Benedetto G, Caramelli D, Secchieri E, Simoni L, Katti E, Malaspina P, Novelletto A, Marin VT, and Barbujani G. 2001. Genetic characterization of the body attributed to the evangelist Luke. *Proc. Natl. Acad. Sci. U.S.A* 98:13460-13463.

TABLE S4

	Posterior distribution population sizes					
	TPM		SM		SDGM	
	UP-Ne	N-Ne	UP-Ne	N-Ne	UP-Ne	N-Ne
Weighted 2.5 % Perc.	187,73	203,71	313,67	359,12	101,05	205,42
Weighted Median	2376,39	1207,48	2748,90	6828,98	2453,44	2885,99
Weighted 97.5 % Perc.	4897,56	64582,69	5094,00	78293,07	5145,36	86737,85

Posterior distribution of Upper Palaeolithic and Neolithic effective population sizes, estimated under an ABC approach.

	Power to recover the true model		
	% of model attribution (a)		
	TPM	SM	SDGM
TPM	60,20	15,06	24,20
SM	45,40	26,00	28,60
SDGM	31,40	24,80	43,80

Power to recover the true model randomly sampling 500 replicates within the 1 million simulated dataset under the TPM (first row), the SM (second row) and the SDGM (third row). The posterior probabilities were all obtained using the neuralnet method with a $\text{tol}=0.01$, as implemented in the R `abc` package. (a) Proportion of times that the method recovered the true model. Replicates were considered assigned to the model that has the highest posterior distribution.

TABLE S5

Mitochondrial DNA profiles of the laboratory staff, archaeologists and anthropologists		
Code	Staff	HVI sequence
Lab_Madrid_1*	Laboratory staff (handling samples in all stages)	16126C, 16294T, 16296T, 16304C
Lab_Madrid_2	Laboratory staff (handling samples only in post-PCR stages)	16224C, 16311C
Lab_Madrid_3	Laboratory staff (having access to the pre-PCR and PCR rooms)	CRS
Lab_Madrid_5	Laboratory staff (having access to the PCR room)	16069T, 16126C
Lab_Bordeaux_1*	Laboratory staff (handling samples in all stages)	16126C, 16294T, 16296T, 16304C
Lab_Bordeaux_2	Laboratory staff (having access to the pre-PCR and PCR rooms)	16126C, 16294T, 16324C
Lab_Bordeaux_3	Laboratory staff (having access to the pre-PCR and PCR rooms)	16209C
Lab_Bordeaux_4	Laboratory staff (having access to the pre-PCR and PCR rooms)	CRS
Lab_Bordeaux_5	Laboratory staff (having access to the pre-PCR and PCR rooms)	16234T
Can_Sadurní_1	Archaeologist	16069T, 16126C, 16278T, 16366T
Can_Sadurní_1	Archaeologist	16126C, 16239T, 16294T, 16296T, 16304C
Can_Sadurní_2	Archaeologist	16126C, 16163G, 16186T, 16189C, 16261T, 16294T
Can_Sadurní_3	Anthropologist	16270T
Can_Sadurní_4	Anthropologist	CRS
Can_Sadurní_5	Anthropologist	CRS
Chaves_1	Archaeologist	16293G
Chaves_2	Archaeologist	16037G, 16183C, 16189C
Chaves_3	Anthropologist	16224C, 16311C
Chaves_4	Archaeologist	16069T, 16126C
Sant_Pau_del_Camp_1	Anthropologist	16304C, 16320T
Sant_Pau_del_Camp_2	Anthropologist	16168C
Sant_Pau_del_Camp_3	Anthropologist	16265G
Sant_Pau_del_Camp_4	Anthropologist	16311C

* Asterisk refers to the same person who worked in both labs (CG).

TABLE S6

Primers used for the amplification of HVRI and coding regions of the mtDNA

Primer pair	Start FW	Primer Forward (5'-3')	End FW	Start RV	Primer Reverse (5'-3')	End RV	Reference
HVI_3,2	16106	GCCAGCCACCATGAATATTG	16125	16280	TGGTATCCTAGTGGGTGAG	16262	Fernández, 2005
HVI_3,4	16106	GCCAGCCACCATGAATATTG	16125	16262	ATCCTAGTGGGTGAGGGG	16280	Fernández, 2005
HVI_1,4	16095	CGTACATTACTGCCAGCC	16112	16262	ATCCTAGTGGGTGAGGGG	16280	Fernández, 2005
HVI_5,6	16232	CACACATCAACTGCAACTCC	16251	16399	TCAAGGGACCCCTATCTGAG	16380	Fernández, 2005
HVI_cu,cl	16190	CCCCATGCTTACAAGCAAGT	16209	16410	GAGGATGGTGGTCAAGGGAC	16391	Jeahas et al., 2001
HVI_bu,bl	16144	TGACCACCTGTAGTACATAA	16163	16251	GGAGTTGCAGTTGATGT	16235	Jeahas et al., 2001
1A_18	15975	CTCCACCATTAGCACCCAA	15973	16418	ATTTCACGGAGGATGGTG	16401	Gabriel et al. 2003
SNP_5178	5129	CTTAACTCCAGCACACGA	5148	5238	GGGCAAAAAGCCGGTTAG	16251	Present study
SNP_4833	4790	ATAGCCCCCTTCACTTCTG	4810	4887	GGGAGAGATTTGGTATATGATTGA	4910	Present study
SNP_10873	10825	AATTTGAATCAACACAACCA	10844	10901	GGGGAACAGCTAAATAGGTT	10920	Present study
SNP_10398_10400	10359	AGTCTGGCCTATGAGTGACTAC	10380	10423	AATGAGTCGAAATCATTGTTT	10444	Present study
SNP_10238	10174	GCGGCTTCGACCCTATATC	10192	10279	TTGTAGGGCTCATGGTAGGG	10298	Present study
SNP_10550	10500	GCATTACCATCTCACTTCTAGG	10522	10609	GGAGTGGGTGTTGAGGGTTA	10628	Present study
SNP_7028	6979	CAAACATCACTAGACATCG	6999	7066	GAATGAAGCCTCCTATGATGG	7086	Present study
SNP_14766	14712	AAAACCATCGTTGTATTCAA	14732	14792	GGAGGTCGATGAATGAGTG	14810	Present study
SNP_3197	3116	CCTCCCTGTACGAAAGGACA	3135	3228	GGGCTCTGCCATCTTAACAA	3247	Present study
SNP_1719	1689	CCCACTCCACCTTACTACCAGA	1710	1755	TGCGCCAGGTTTCAATT	1772	Nelson et al., 2007
SNP_6371	6267	GCAGGAACAGGTTGAACAGTC	6287	6377	TGAAATTGATGGCCCTAAG	6396	Present study

FW= Forward

RV=Reverse

References

Fernández E. 2005. Polimorfismos de DNA mitocondrial en poblaciones antiguas de la cuenca mediterránea. Available from: www.tesisenxarxa.net/TESIS_UB/AVAILABLE/TDX-0123106-084234/

Gabriel MN, Calloway CD, Reynolds RL, and Primorac D. 2003. Identification of human remains by immobilized sequence-specific oligonucleotide probe analysis of mtDNA hypervariable regions I and II. *Croat. Med. J* 44:293-298.

Jeahas E, Toprak K, Vanderheyden N, Pfeiffer H, Cassiman JJ, Brinkmann B, and Decorte R. 2001. Pitfalls in the analysis of mitochondrial DNA from ancient specimens and the consequences for forensic DNA analysis: the historical case of the putative heart of Louis XVII. *Int. J. Legal Med* 115:135-141.

Nelson TM, Just RS, Loreille O, Schanfield MS, and Podini D. 2007. Development of a multiplex single base extension assay for mitochondrial DNA haplogroup typing. *Croat. Med. J* 48:460-472.

TABLE S7

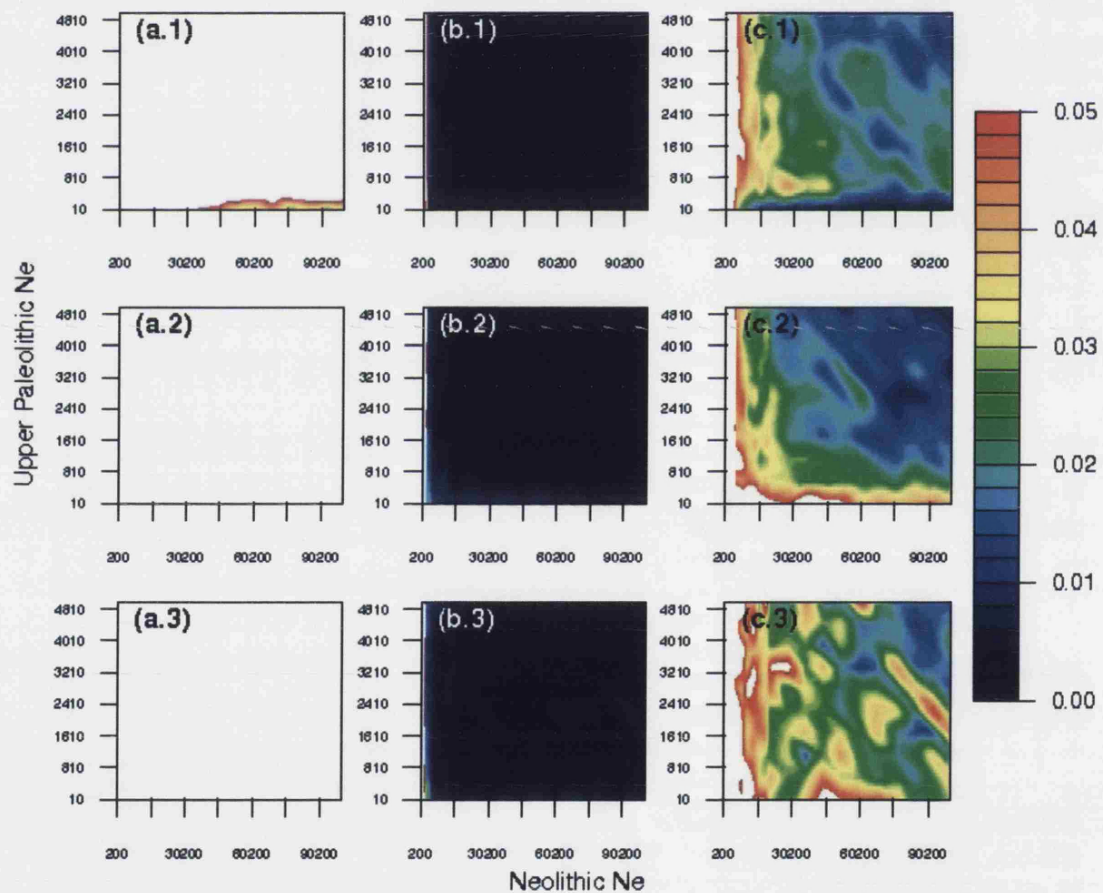
Cloned Sequences alignment

Sequences are identified by sample name, extraction number, amplification number, primer pair and clone number or “dir” (direct sequence). Sequences follow specimen order listed in **Table 1** and **Table S1**. Colours identify discarded haplotypes (grey=lab staff/archaeologist/anthropologist contamination, orange=cross-contamination) and jumping-PCR (blue).

AVAILABLE IN THE **CD** PROVIDED TOGETHER WITH THIS PHD THESIS.

FIGURE S1

Probabilities of obtaining simulated F_{ST} greater than observed are plotted (UP- N_e : Upper Palaeolithic effective population size; N- N_e : Neolithic Effective population size). (a-c.1) TPM model, (a-c.2) SM model, (a-c.3) SDGM model, (a.1-3) MN and modern-day populations F_{ST} ; (b.1-3) EN and modern-day populations F_{ST} ; (c.1-3) MN and EN populations F_{ST} . Values higher than 0.05 are in white. Include all N- N_e values explored (up to 100,000).



ARTÍCULO 4 / ARTICLE 4

Incluye una actualización con los resultados validados y el análisis estadístico y computacional / Update with the results validated and the statistical and computational analyses included

Population genetics and DNA preservation in ancient human remains from Eastern Spain.

Gamba C, Fernández E, Oliver A, Tirado M, Baeza C, López-Parra A, y Arroyo-Pardo E.

Forensic Science International: Genetics Supplement Series 1:462-464. 2008.

RESUMEN (ESPAÑOL)

El objetivo de este trabajo era evaluar el impacto genético de diferentes pobladores que habitaron el noreste de la Península Ibérica desde el Calcolítico hasta la Edad del Hierro. Más concretamente, se evaluó la continuidad o discontinuidad genética entre los diferentes grupos estudiados y las poblaciones neolíticas y actuales de la misma región.

Se obtuvieron resultados validados para 12 individuos, 8 de la Edad de los Metales (Calcolítico y Edad del Bronce) y 4 Íberos (Edad del Hierro). Se implementaron diferentes modelos estadísticos y computacionales para la evaluación del impacto genético de los diferentes pobladores de la región. Los resultados apuntan a la presencia de cierta continuidad genética desde el Neolítico Medio hasta la actualidad, a excepción de la población íbera. Esta última presentaba una composición genética diferencial que podría atribuirse a diferentes factores: al escaso número de muestras estudiadas, a una elevada endogamia o a que se tratase de un grupo étnico diferencial, puesto que los individuos analizados no fueron incinerados como corresponde a la tradición funeraria de ese período.

En el conjunto de datos analizados destaca la presencia de haplogrupos mitocondriales tradicionalmente asociados a la expansión post-glacial (como el V y el HV0) desde los refugios paleolíticos del sur, de acuerdo con los datos de la genética de poblaciones actuales. Su presencia en las muestras estudiadas apunta a una probable localización más meridional de estos refugios durante el último máximo glacial.

Research article

Population genetics and DNA preservation in ancient human remains from Eastern Spain

C. Gamba^a, E. Fernández^a, A. Oliver^b, M. Tirado^a, C. Baeza^a,
A.M. López-Parra^a, E. Arroyo-Pardo^{a,*}^a *Depto. Toxicología y Legislación Sanitaria, Facultad de Medicina, Universidad Complutense, Madrid 28040, Spain*^b *Sección de Arqueología, Museo de Bellas Artes de Castellón, Castellón, Spain*

Received 16 August 2007; accepted 7 October 2007

Abstract

This work aims to establish the genetic relationship between the different settlers of Eastern Spain and also to determine the conditions of DNA preservation. We studied two overlapping sequences (16,126–16,251 and 16,256–16,369) from mitochondrial HVR-I in 37 bone and teeth samples from 17 archaeological sites of Spanish Levant. Consistence of the results was established by repeated replication of amplifications. Approximately, 50% of the samples yielded reproducible results. The high efficiency in DNA recovery indicates that sample preservation mainly depends on the depositional environment rather than on sample age. Haplogroup V, an alleged marker of Palaeolithic newcomers in Europe, has been found in an unusual elevated frequency (1 Chalcolithic and 2 Iberian samples). This result could suggest a more southern distribution of Palaeolithic ice refugia. Moreover, we found Haplogroup L in Chalcolithic samples. This may suggest the presence of a prehistoric African genetic background in eastern Iberia.

© 2008 Elsevier Ireland Ltd. All rights reserved.

Keywords: Ancient DNA; Mitochondrial DNA; Eastern Iberia

1. Introduction

The East of Spain has been populated at least since Palaeolithic times [1]. Due to its strategic location, opened to the Mediterranean shore, it has been prone to have received genetic influences from other populations. Present work tries to evaluate this hypothesis through mtDNA Hypervariable region I (HVR-I) characterization of ancient samples from Castellón and Valencia provinces, dated back to Roman, Iberian, Bronze and Chalcolithic ages. Obtained results will be interpreted in terms of genetic continuity/discontinuity between these periods. The effect of the environment and the age of the sample over DNA preservation will be evaluated in this study.

2. Material and methods

2.1. Samples

36 samples were studied (21 bones and 15 teeth) belonging to 36 different individuals from 17 archaeological sites (6

Chalcolithic, 2 Bronze Age, 6 Iberian Culture and 3 Roman) located in Eastern Spain (see Fig. 1 and Table 1).

2.2. Extraction

The external surface of the samples was removed first with a Sand-blasted, and afterwards they were irradiated with UV light and crushed in a freezer mill filled with liquid Nitrogen. Samples were then washed several times with EDTA 0.5 M and incubated in a lysis solution (EDTA 5 mM, Tris-HCl 10 mM, SDS 0.5%, proteinase K 50 µg/ml) overnight at 36 °C. DNA was extracted using a modified Phenol-Chloroform protocol [2] and concentrated with microcons (Centriplus 30000, Millipore).

2.3. PCR amplification and sequencing

A 244-bp fragment was studied from mitochondrial HVRI by the independent amplification of two short overlapping sequences (positions 16,126–16,369) [2]. Qiagen Multiplex PCR was implemented for ancient DNA. Amplification products were sequenced using a Applied Biosystems 310 sequencer.

* Corresponding author. Tel.: +34 91 3941576; fax: +34 91 3941606.

E-mail address: earroyop@med.ucm.es (E. Arroyo-Pardo).

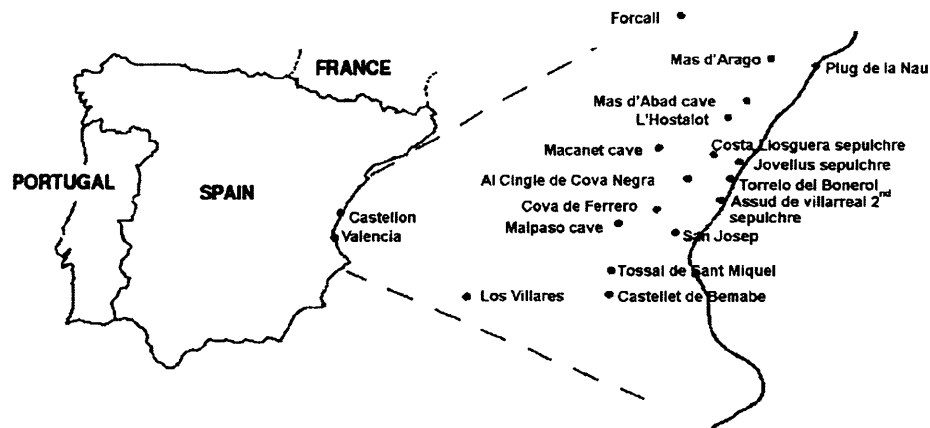


Fig. 1. Location of sample sites.

Table 1
Analysed samples

Archaeological site	Location	No.	Sample	Age
Malpaso cave	Castellónovo (Castellón)	4	Teeth	Chalcolithic
Jovellus sepulchre	Benicàssim (Castellón)	1	Tooth	Chalcolithic
Al Cingle de Cova Negra	Borriol (Castellón)	1	Tooth	Chalcolithic
Ferrero cave	Artana (Castellón)	1	Bone	Chalcolithic
Assud de Villarreal, 2nd sepulchre	Almassora (Castellón)	1	Bone	Chalcolithic
Costa Lloguera sepulchre	Castellón	6	Teeth	Chalcolithic
Mas d'Abad cave	Les Coves de Vinromà (Castellón)	1	Bone	Bronze Age
Maçanet cave	Llucena (Castellón)	1	Tooth	Bronze Age
Castellet de Bernabé	Liria (Valencia)	2	Bones	Iberian
Los Villares	Caudete de las Fuentes (Castellón)	1	Bone	Iberian
Puig de la Nau	Benicarló (Castellón)	9	Bones	Iberian
Torreló Bonerol	Castellón	2	Bones	Iberian
Sant Josep	La Vall d'Uixó (Castellón)	1	Bone	Iberian
Tossal de Sant Miquel	Liria (Valencia)	3	Bones	Iberian
Mas d'Aragó	Cervera del Maestre (Castellón)	1	Tooth	Roman
L'Hostalot	Villanova d'Alcolea (Castellón)	1	Tooth	Roman
Total		36		

2.4. Criteria of authenticity

Some criteria of authenticity were followed in this study: analysis in an exclusive ancient DNA lab, physical separation of pre- and post-PCR procedures, equipment exclusive for aDNA, removal of the external surface of the samples analysis by a single researcher, use of “blanks” in extraction and amplification, repeated amplifications from the same extract, phylogenetic sense (comparison with lab staff and archaeologists/anthropologists), DNA quantification.

3. Results

Approximately, 50% of the samples yielded reproducible results: 17 samples for the two fragments and 3 samples only for one of them (See Tables 1 and 2).

By comparing the obtained haplotypes with public databases it was possible to assign an haplogroup to most of the sequences. Haplogroup H, the most common nowadays in Europe, was found at all age classes. Haplogroup V, an alleged

marker for human post-glacial expansion from southwestern refugia [3–4] has been found in an unusual elevated frequency (1 Chalcolithic and 3 Iberian samples: 12.5% and 33.3%, respectively). Moreover, the Chalcolithic period is characterized by the possible presence of haplogroup L3, a typical sub-saharian haplogroup [5].

4. Discussion

The high efficiency in DNA recovery obtained indicates that sample preservation mainly depends on the depositional environment rather than on sample age. Moreover, cave environment seems to aid DNA preservation as it maintains temperature in a constant interval over years. Thermal history has been proved to play an important role in the preservation of other biomolecules as amino acids [6]. The unique haplotype shared by individuals from different chronologies is *CRS*. However, as this is the more frequent haplotype nowadays, its presence alone does not allow us to infer the existence of genetic continuity through those periods. Despite the reduced

Table 2
Obtained results

Sample	Arch. site	Period	Haplotype	Haplogroup
1MP1	Malpaso cave	Chalcolithic	16223T, 16325C, 16362C	D
4MP12	Malpaso cave	Chalcolithic	16298C*	V
2MP4	Malpaso cave	Chalcolithic	16218T, 16223T	L3?
1JO1	Jovellús sepulchre	Chalcolithic	16218Y, 16223Y	L3?
1AV1	Assud de Villareal, 2nd sepulchre	Chalcolithic	16218Y, 16223Y	L3?
COST1	Costa Lloguera sepulchre	Chalcolithic	16293R*	H (60%)
COST3	Costa Lloguera sepulchre	Chalcolithic	16223Y	L3a?
COST4	Costa Lloguera sepulchre	Chalcolithic	CRS	H
1MN1	Maçanet cave	Bronze Age	CRS	H
CB2	Castellet de Bernabé	Iberian	16163G, 16186T, 16189C, 16250Y, 16294T, 16301Y	T1
LV1	Los Villares	Iberian	16298C	V
PB14	Puig de la Nau	Iberian	16293T, 16298C	V
PB18	Puig de la Nau	Iberian	CRS	H
PB19	Puig de la Nau	Iberian	CRS	H
PB20	Puig de la Nau	Iberian	16293T, 16298C	V
TORR1	Torreló Bonerot	Iberian	16223T, 16292Y, 16295Y, 16304Y, 16311Y	W
TORR2	Torreló Bonerot	Iberian	16223T, 16292Y, 16295Y, 16304Y	W
TSM1	Tossal de Sant Miquel	Iberian	CRS*	H?
TSM2	Tossal de Sant Miquel	Iberian	16224Y, 16234Y, 16311Y	K
1MAR2	Mas d'Aragó	Roman	CRS	H

* Indicates the failure of the recovery of the first fragment.

amount of samples, some interesting conclusions regarding to the ancient Iberian genetic background could be drawn. Haplogroup V is considered to have originated at the Cantabrian fringe during the last Ice Age, from which it is thought to have expanded northwards after the ice retreat ~10,000–15,000 YBP [3–4]. However, ancient DNA analysis failed to detect this haplogroup in ancient Basque population [7]. Its presence in Eastern Spain in Chalcolithic and Iberian periods could be explained as a result of prehistoric migrations from the north, but also this fact could be indicating a more southern distribution of Palaeolithic ice refugia. The present work also suggests the presence of a prehistoric African genetic background in eastern Iberia. The existence of genetic African lineages in prehistory has been ascertained in other North Iberian populations also in Chalcolithic period [2–8]. Together these results point out at a pre-Chalcolithic migration from Africa into Iberia with subsequent genetic replacement.

Acknowledgments

This work has been funded by the Research Project CGL2006-07828/BOS of the Spanish Ministry of Education and by the Dávalos Fletcher Foundation—Fine Arts Museum of Castellón.

Conflict of interest

None.

References

- [1] I. Barandiarán et al., Prehistoria de la Península Ibérica, Ariel Prehistoria, 4th edition, 2004.
- [2] E. Fernández, Polimorfismos de DNA mitocondrial en poblaciones antiguas de la Cuenca Mediterránea, Ph.D. Thesis, University of Barcelona, 2005.
- [3] A. Torroni, et al., MtDNA analysis reveals a major late Paleolithic population expansion from south-western to north-eastern Europe, *Am. J. Hum. Genet.* 62 (1998) 1137–1152.
- [4] A. Torroni, et al., A signal, from human mtDNA, of postglacial recolonization in Europe, *Am. J. Hum. Genet.* 69 (4) (2001) 844–852.
- [5] A. Salas, et al., The making of the African mtDNA landscape, *Am. J. Hum. Genet.* 71 (5) (2002) 1082–1111.
- [6] T. Torres, et al., Bear dentine aspartic acid racemization analysis: a proxy for the dating of Pleistocene cave infills, *Archaeometry* 44 (3) (2002) 417–426.
- [7] N. Izaguirre, C. de la Rúa, A mtDNA analysis in ancient Basque populations: implications for haplogroup V as a marker for a major paleolithic expansion from southwestern Europe, *Am. J. Hum. Genet.* 65 (1) (1999) 199–207.
- [8] A. Alzualde, et al., Insights into the “isolation” of the Basques: Mtdna, *Am. J. Phys. Anthropol.* 130 (3) (2006) 394–404.

ACTUALIZACIÓN - ARTÍCULO 4

El conjunto de muestras analizadas en el ARTÍCULO 4 fue posteriormente ampliado, aumentando el número de individuos de cada período. Además, los resultados obtenidos fueron analizados según los criterios de autenticidad y el enfoque computacional descritos en el ARTÍCULO 3.

Se adjuntan los siguientes anexos con los resultados genéticos y computacionales:

- **Tabla A1:** Haplotipos y haplogrupos validados.
- **Tabla A2:** Distancias genéticas (F_{ST}) entre las poblaciones estudiadas.
- **Anexo I:** Figuras de las comparaciones entre F_{ST} observadas y simuladas.
- **Anexo II:** Alineamiento de las secuencias clonadas.

TABLAS

Tabla A1. Haplotipos y haplogrupos mitocondriales.

Yacimiento	Período	Datación		Individuo	Haplotipo	SNPs tipados	Haplogrupo
Tossal Sant Miquel	Íberos (Edad del Hierro)	del	Siglo IV-III a.C.	TSM2	16311C	10550G	K
Los Villares	Íberos (Edad del Hierro)	del	Siglo IV-III a.C.	LV1	16298C	4580G	HV0
Puig de la Nau	Íberos (Edad del Hierro)	del	Siglo V a.C.	PB14	16293T 16298C	4580G	HV0
Puig de la Nau	Íberos (Edad del Hierro)	del	Siglo V a.C.	PB20	16293T 16298C	4580G	HV0
Maçanet	Edad de los Metales (Bronze Age)		1.500-1000 a.C.	1MN	16069T 16126C	13708A	J
Mas d'Abad	Edad de los Metales (Bronze Age)		1.500-1000 a.C.	3MA	16189C 16224C 16311C	10550G	K
Sepulcre Segond de la Costa	Edad de los Metales (Chalcolithic)		2500-2000 a.C.	COST1	16136C 16192T 16270T 16304C	3197C 7028T 10398A 10873T 10400C	U53b
Sepulcre Segond de la Costa	Edad de los Metales (Chalcolithic)		2500-2000 a.C.	COST4	CRS	4580G 7028T 14766C	HV
Torre d'en Domenèch	Edad de los Metales (Calcolítico)		2500-2000 a.C.	1TD	16224C 16311C	10550G	K
Cova de Malpaso	Edad de los Metales (Calcolítico)		2500-2000 a.C.	1MP	16223T 16362C	10873T 10400C 14766T	N*
Racó de Raca	Edad de los Metales (Calcolítico)		2500-2000 a.C.	RR1	16189C 16223T 16278T	6371T 10873T 14766T	X2
Assud de Villarreal	Edad de los Metales (Calcolítico)		2500-2000 a.C.	1AV	16298C	4580A	V

Tabla A2. Distancias genéticas (F_{ST}) entre las poblaciones analizadas.

Edad de la Población	Moderna	Íbera	Edad de los Metales	Neolítica Media	Neolítica Antigua
Moderna					
Íbera	0,18525				
Edad de los Metales	0,02232	0,10307			
Neolítica Media	0,23213	0,18178	0,01261		
Neolítica Antigua	0,12978	0,22252	-0,02577	0,10417	

Figura A1. Probabilidad de obtener F_{ST} simulados mayores que los F_{ST} observados entre la población moderna (García et al., 2011) y la de la Íbera (ARTÍCULO 4), en todos los rangos de N_e explorados y en correspondencia de los tres modelos (TMP, SM, SDGM).

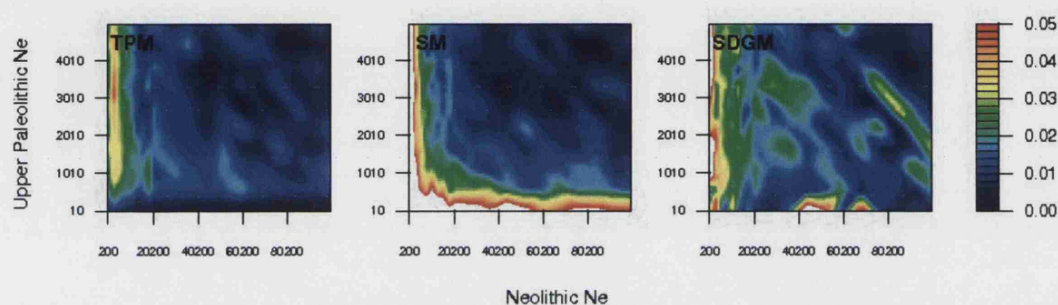


Figura A2. Probabilidad de obtener F_{ST} simulados mayores que los F_{ST} observados entre la población moderna (García et al., 2011) y la de la Edad de los Metales (ARTÍCULO 4), en todos los rangos de N_e explorados y en correspondencia de los tres modelos (TMP, SM, SDGM).

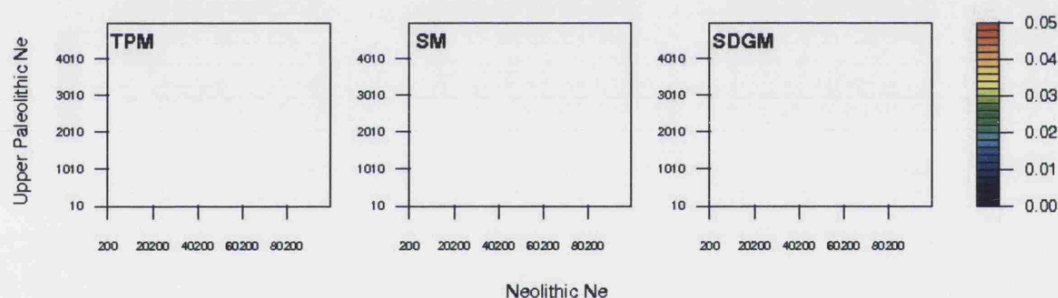


Figura A3. Probabilidad de obtener F_{ST} simulados mayores que los F_{ST} observados entre la población Íbera (ARTÍCULO 1)(Gamba et al., 2008)(Gamba et al., 2008)(Gamba et al., 2008)(Gamba et al., 2008) y la de la Edad de los Metales (ARTÍCULO 4), en todos los rangos de N_e explorados y en correspondencia de los tres modelos (TMP, SM, SDGM).

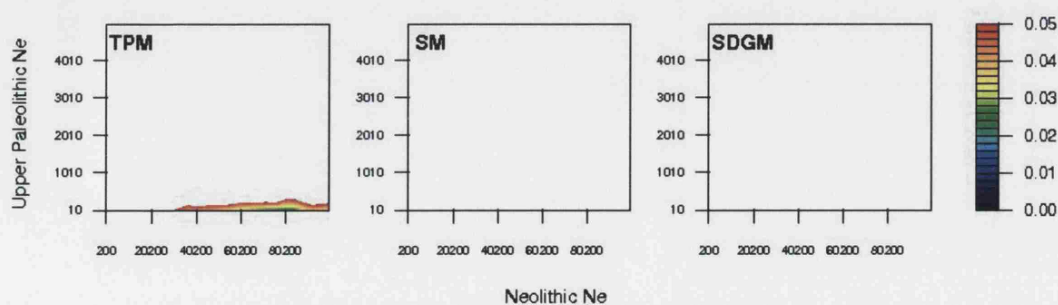


Figura A4. Probabilidad de obtener F_{ST} simulados mayores que los F_{ST} observados entre la población Neolítica Media (Sampietro et al., 2007) y la Íbera (ARTÍCULO 4), en todos los rangos de N_e explorados y en correspondencia de los tres modelos (TMP, SM, SDGM).

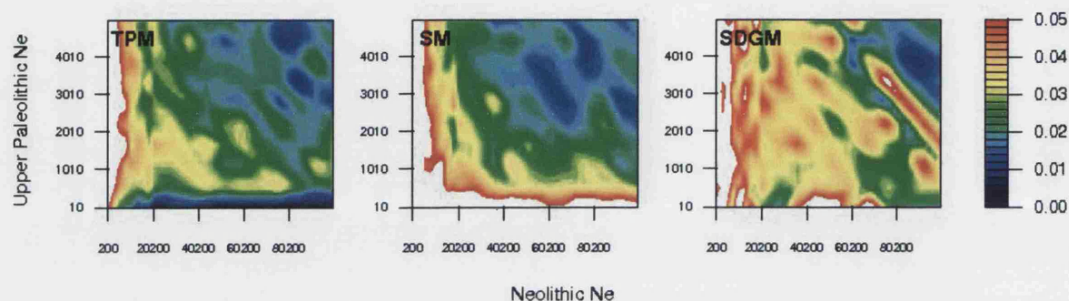


Figura A5. Probabilidad de obtener F_{ST} simulados mayores que los F_{ST} observados entre la población Neolítica Antigua (ARTÍCULO5) y la Íbera (ARTÍCULO 4) en todos los rangos de N_e explorados y en correspondencia de los tres modelos (TMP, SM, SDGM).

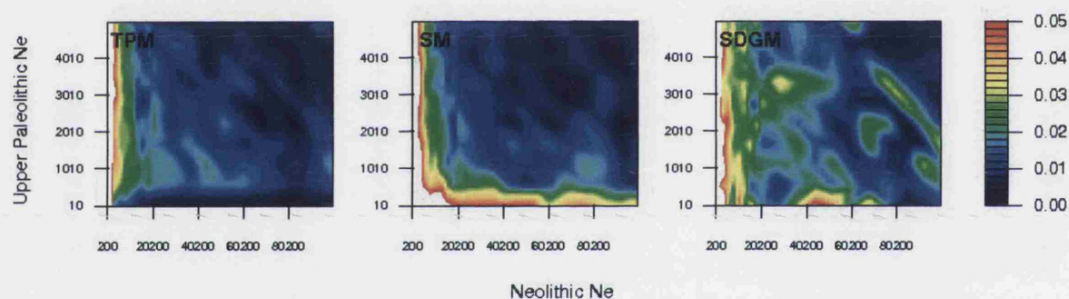


Figura A6. Probabilidad de obtener F_{ST} simulados mayores que los F_{ST} observados entre la población Neolítica Media (Sampietro et al., 2007) y la de la Edad de los Metales (ARTÍCULO 4), en todos los rangos de N_e explorados y en correspondencia de los tres modelos (TMP, SM, SDGM).

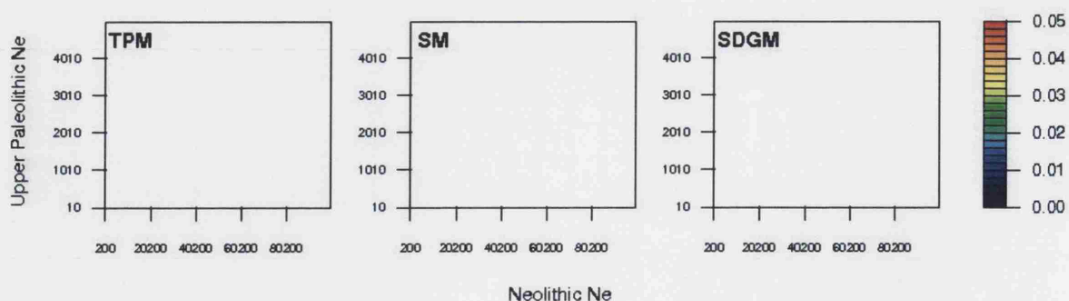
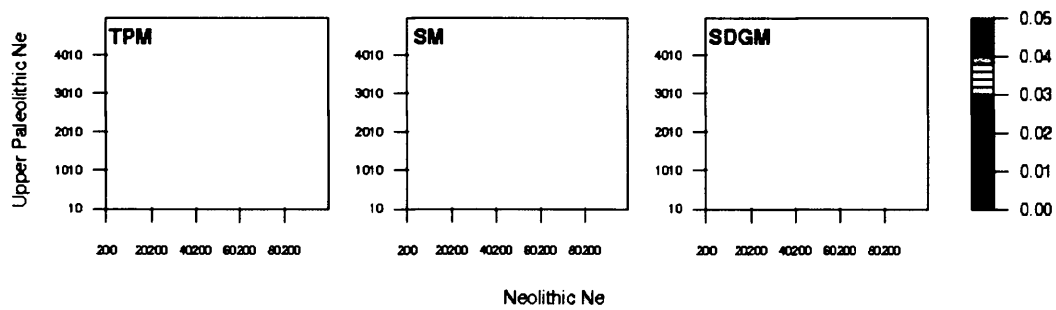


Figura A7. Probabilidad de obtener F_{ST} simulados mayores que los F_{ST} observados entre la población Neolítica Antigua (ARTÍCULO 5) y la de la Edad de los Metales (ARTÍCULO 4), en todos los rangos de N_e explorados y en correspondencia de los tres modelos (TMP, SM, SDGM).



Alineamiento de las secuencias clonadas

Las secuencias están identificadas por el nombre de la muestra, el número de extracción, el de amplificación, el par de cebadores y el número del clon o la sigla “dir” (secuencia directa). Las secuencias siguen el mismo orden de individuos listados en las Tablas 1 y S1. Los colores identifican los haplotipos descartados (gris = contaminación por parte del personal del laboratorio/arqueólogos/antropólogos, naranja = contaminación entre muestras) y en azul se destaca el fenómeno de *Jumping PCR*.

DISPONIBLE EN EL **CD** PROPORCIONADO JUNTO CON ESTA TESIS DOCTORAL.

Sample size analysed in the ARTICLE 4 has been extended, increasing the number of individual per each period. Moreover, the results obtained have been analysed following the criteria of authenticity and the computational approach described in the ARTICLE 3.

Genetic and computational results are included in the following Appendix:

- **Table A1:** Validated haplotypes and haplogroups.
- **Table A2:** Genetic distances (F_{ST}) among the populations compared.
- **Appendix I:** Comparison between observed and simulated F_{ST} .
- **Appendix II:** Cloned sequences alignment.

TABLES

Table A1. Mitochondrial haplotypes and haplogroups.

Site	Period	Date	Individual	Haplotype	SNPs typed	Haplogroup
Tossal Sant Miquel	Iberic (Iron Age)	IV-III Centuries B.C.	TSM2	16311C	10550G	K
Los Villares	Iberic (Iron Age)	IV-III Centuries B.C.	LV1	16298C	4580G	HV0
Puig de la Nau	Iberic (Iron Age)	V Century B.C.	PB14	16293T 16298C	4580G	HV0
Puig de la Nau	Iberic (Iron Age)	V Century B.C.	PB20	16293T 16298C	4580G	HV0
Maçanet	Metal Age (Bronze Age)	1.500-1000 B.C.	1MN	16069T 16126C	13708A	J
Mas d'Abad	Metal Age (Bronze Age)	1.500-1000 B.C.	3MA	16189C 16224C 16311C	10550G	K
Sepulcre Segond de la Costa	Metal Age (Chalcolithic)	2500-2000 B.C.	COST1	16136C 16192T 16270T 16304C	3197C 7028T 10398A 10873T 10400C	U53b
Sepulcre Segond de la Costa	Metal Age (Chalcolithic)	2500-2000 B.C.	COST4	CRS	4580G 7028T 14766C	HV
Torre d'en Domenèch	Metal Age (Chalcolithic)	2500-2000 B.C.	1TD	16224C 16311C	10550G	K
Cova de Malpaso	Metal Age (Chalcolithic)	2500-2000 B.C.	1MP	16223T 16362C	10873T 10400C 14766T	N*
Racó de Raca	Metal Age (Chalcolithic)	2500-2000 B.C.	RR1	16189C 16223T 16278T	6371T 10873T 14766T	X2
Assud de Villarreal	Metal Age (Chalcolithic)	2500-2000 B.C.	1AV	16298C	4580A	V

Table A2. Genetic distances (F_{ST}) among the populations compared.

Age of the population	Modern	Iberic	Metal Age	Middle Neolithic	Early Neolithic
Modern					
Iberic	0,18525				
Metal Age	0,02232	0,10307			
Middle Neolithic	0,23213	0,18178	0,01261		
Early Neolithic	0,12978	0,22252	-0,02577	0,10417	

Figure A1. Probability of obtaining simulated F_{ST} greater than observed between modern (García et al., 2011) and Iberic populations (ARTICLE 4), for all N_e ranges explored and in correspondence of the three demographic models tested (TMP, SM, SDGM). Values higher than 0.05 are in white.

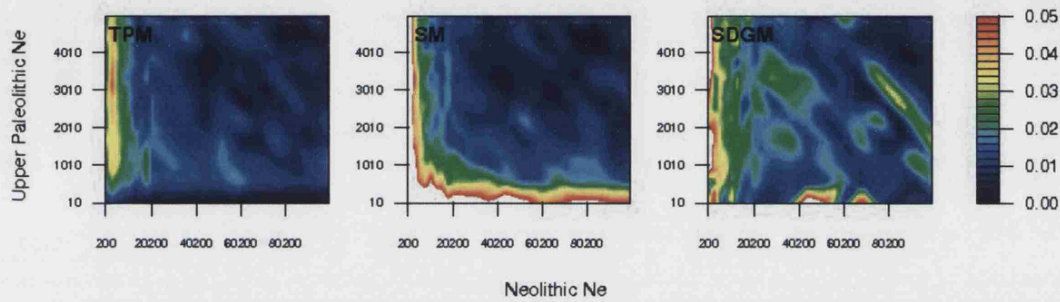


Figure A2. Probability of obtaining simulated F_{ST} greater than observed between modern (García et al., 2011) and Metal Age populations (ARTICLE 4), for all N_e ranges explored and in correspondence of the three demographic models tested (TMP, SM, SDGM). Values higher than 0.05 are in white.

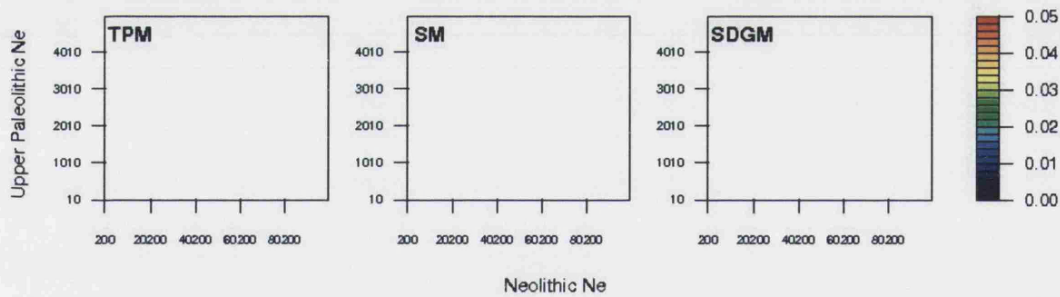


Figure A3. Probability of obtaining simulated F_{ST} greater than observed between Iberic and Metal Age populations (ARTICLE 4), for all N_e ranges explored and in correspondence of the three demographic models tested (TMP, SM, SDGM). Values higher than 0.05 are in white.

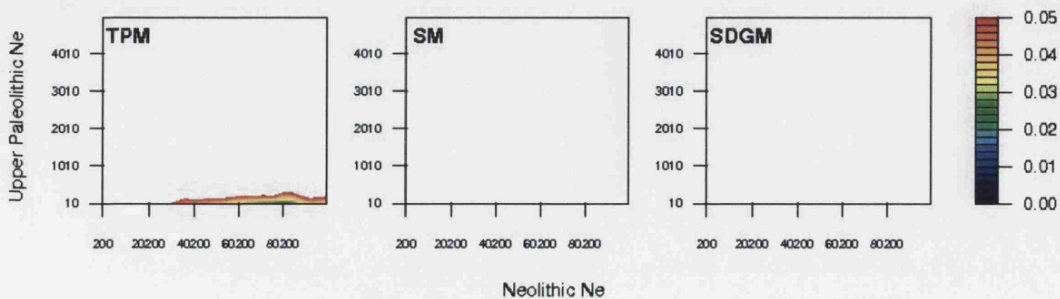


Figure A4. Probability of obtaining simulated F_{ST} greater than observed between Middle Neolithic (Sampietro et al., 2007) and Iberic populations (**ARTICLE 4**), for all N_e ranges explored and in correspondence of the three demographic models tested (TPM, SM, SDGM). Values higher than 0.05 are in white.

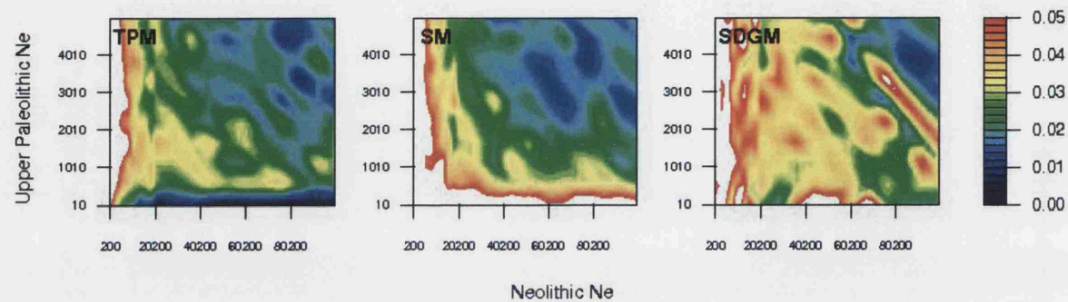


Figure A5. Probability of obtaining simulated F_{ST} greater than observed between Early Neolithic (**ARTICLE 3**) and Iberic populations (**ARTICLE 4**), for all N_e ranges explored and in correspondence of the three demographic models tested (TPM, SM, SDGM). Values higher than 0.05 are in white.

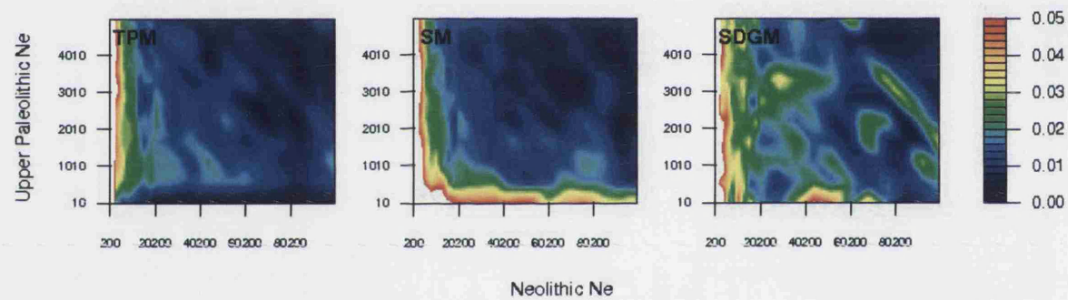


Figure A6. Probability of obtaining simulated F_{ST} greater than observed between Middle Neolithic (Sampietro et al., 2007) and Metal Age populations (**ARTICLE 4**), for all N_e ranges explored and in correspondence of the three demographic models tested (TPM, SM, SDGM). Values higher than 0.05 are in white.

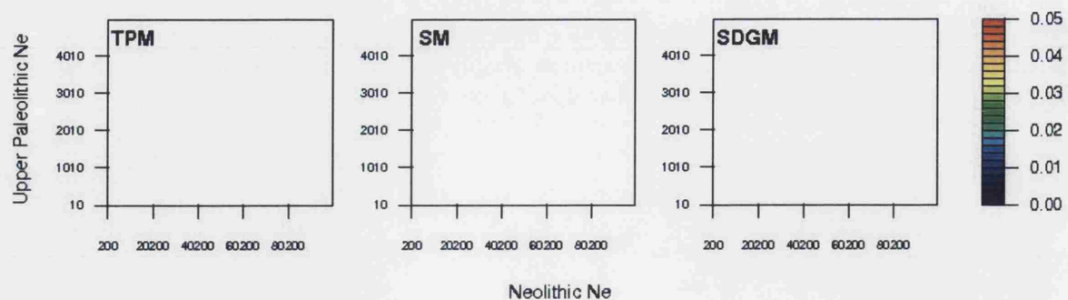
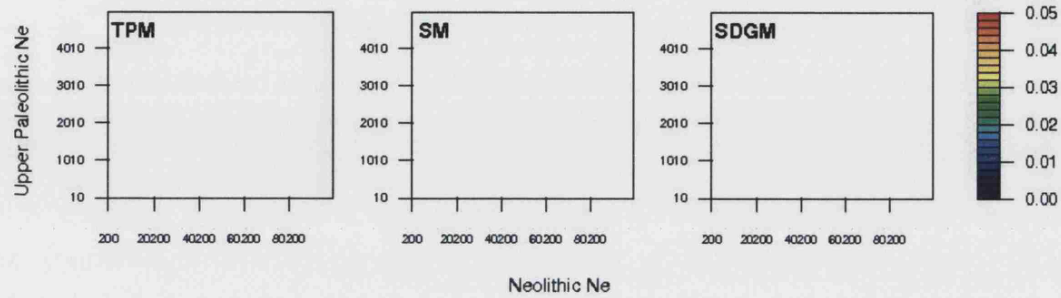


Figure A7. Probability of obtaining simulated F_{ST} greater than observed between Early Neolithic (**ARTICLE 3**) and Metal Age populations (**ARTICLE 4**), for all N_e ranges explored and in correspondence of the three demographic models tested (TPM, SM, SDGM). Values higher than 0.05 are in white.



Cloned sequences alignments

Sequences are identified by sample name, extraction number, amplification number, primer pair and clone number or “dir” (direct sequence). Sequences follow specimen order listed in Table 1 and Table S1. Colours identify discarded haplotypes (grey=lab staff/archaeologist/anthropologist contamination, orange=cross-contamination) and jumping-PCR (blue).

AVAILABLE IN THE **CD** PROVIDED TOGETHER WITH THIS PHD THESIS.

ARTÍCULO 5 / ARTICLE 5

Incluye Material Suplementario / Supplementary Material included

Brief communication: Ancient nuclear DNA and kinship analysis: The case of a medieval burial in San Esteban Church in Cuéllar (Segovia, Central Spain).

Gamba C, Fernández E, Tirado M, Pastor F, y Arroyo-Pardo E.

American Journal of Physical Anthropology 144:485-491. 2011.

RESUMEN (ESPAÑOL)

Este trabajo se centra en el análisis genético de un conjunto de muestras medievales, con el fin de identificar las posibles relaciones familiares entre los individuos estudiados y determinar el sexo mediante la amplificación de marcadores nucleares. Se trata de una de las principales aplicaciones en el campo del ADN antiguo, con un trasfondo multidisciplinar que engloba datos arqueológicos, historiográficos, antropológicos, genealógicos y genéticos.

Se analizó un conjunto de siete individuos medievales muy bien preservados y en algunos casos momificados naturalmente enterrados en cuatro sepulcros de la Iglesia de San Esteban de Cuéllar (Segovia, España). El conjunto incluía dos individuos infantiles y cinco adultos.

Fue posible recuperar información genética a nivel del ADN nuclear para todos los individuos estudiados mediante el empleo del kit MiniFiler. Cuatro especímenes adultos proporcionaron perfiles nucleares completos. Los resultados obtenidos permitieron establecer el sexo de todos los individuos estudiados, excluir ciertas sospechas arqueológicas de relaciones familiares y asociar otras a probabilidades de entre el 68 % hasta el 99.98 %. En su conjunto, se trata de resultados muy satisfactorios desde el punto de vista de la elevada eficiencia de los análisis genéticos y también desde la perspectiva histórico-arqueológica, proporcionando nuevas evidencias a favor de hipótesis formuladas desde otras disciplinas.

Brief Communication: Ancient Nuclear DNA and Kinship Analysis: The Case of a Medieval Burial in San Esteban Church in Cuellar (Segovia, Central Spain)

Cristina Gamba,^{1*} Eva Fernández,¹ Mirian Tirado,¹ Francisco Pastor,² and Eduardo Arroyo-Pardo¹

¹Laboratory of Forensic and Population Genetics, Toxicology and Health Legislation Department, Complutense University of Madrid, Madrid, Spain

²Anatomy Department, Faculty of Medicine, University of Valladolid, Valladolid, Spain

KEY WORDS ancient DNA; mini STRs; human nuclear DNA; molecular kinship; Spain

ABSTRACT The aim of this work was to investigate a very common situation in the archaeological and anthropological context: the study of a burial site containing several individuals, probably related genetically, using ancient DNA techniques. We used available ancient DNA and forensic protocols to obtain reliable results on archaeological material. The results also enabled molecular sex determination to be compared with osteological data. Specifically, a modified ancient DNA extraction method combined with the amplification of nuclear markers with the AmpFISTR[®] MiniFiler[™] kit (Applied Biosystems) was used. Seven medieval individuals buried in four niches dated in the 15th Century

at San Esteban Church in Cuellar (Segovia, Central Spain) were analyzed by the proposed method, and four of seven provided complete autosomal short tandem repeat (STRs) profiles. Kinship analyses comprising paternity and sibship relations were carried out with pedigree-specific software used in forensic casework. A 99.98% paternity probability was established between two individuals, although lower percentages (68%) were obtained in other cases, and some hypothetical kinship relations were excluded. The overall results could eventually provide evidence for reconstructing the historical record. *Am J Phys Anthropol* 144:485–491, 2011. ©2010 Wiley-Liss, Inc.

Less than three decades ago, the first aDNA studies were published (Higuchi et al., 1984; Pääbo 1985, 1986). Recent years have seen an expansion of this field, due mostly to technical advances in specific protocols.

Difficulties in human aDNA analysis are mainly the low copy number (LCN) of amplifiable DNA, its fragmentation and damage, the presence of polymerase chain reaction (PCR) inhibitors, and susceptibility to contamination from modern DNA. Some of these features are also characteristic of DNA recovered in forensic casework. As a result, a set of authenticity criteria were established early to validate aDNA results (Pääbo et al., 2004).

Mitochondrial DNA (mtDNA) has been the preferred target in aDNA studies because of its higher copy number compared with nuclear DNA (nuDNA) (Wiesner et al., 1992). However, in the last few years, reproducible results on ancient nuDNA have been also achieved (Keyser-Tracqui et al., 2003; Ricaut et al., 2005a,b; Hawass et al., 2010). Moreover, almost the entire genome of a 4,000-year-old human specimen preserved in permafrost has been recovered (Rasmussen et al., 2010). However, it has been suggested that dry environments are also suitable for preservation of nuclear DNA sequences (Poinar et al., 2003). Low levels of humidity could favor mummification phenomena and retard hydrolytic degradation of nucleic acids such as depurination and deamination.

Currently, several protocols are available for LCN and degraded DNA recovered in archaeological and forensic contexts. Specific aDNA extraction methods have been published (Loreille et al., 2007; Rohland and Hofreiter 2007a; Anderung et al., 2008), and new forensic amplification kits for nuDNA, such as AmpFISTR[®] MiniFiler[™] (Mulero et al., 2008) and PowerPlex[®] ESX and ESI

Systems (Sprecher et al., 2009) represent interesting tools both for ancient and forensic DNA studies.

Present day archaeogenetic analyses, combined with archaeological and anthropological data, could answer specific questions about historical or prehistorical hypotheses. As in forensic casework, genetic analyses of human archaeological samples may provide a clue to determining sex (Hummel et al., 2000), investigate kinship between burial groups (Schultes et al., 2000; Keyser-Tracqui et al., 2003; Amory et al., 2007; Hawass et al., 2010), or identify historical figures (Gill et al.,

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: Fundación del Patrimonio Histórico de Castile and León (Government of Castile and León, Spain); Grant sponsor: Ministry of Science and Innovation (MICINN) of the Spanish Government; Grant numbers: CGL2006-07828/BOS and CGL2009-07959; Grant sponsor: PhD FPU; Grant number: AP2006 01586; Grant sponsor: MICINN (to C.G.); Grant sponsor: Post-doctoral research contract “Juan de La Cierva”; Grant sponsor: MICINN and European Social Fund (European Union) (to E.F.).

*Correspondence to: Cristina Gamba, Laboratory of Forensic and Population Genetics, Toxicology and Health Legislation Department, Complutense University of Madrid, Madrid, Spain 28040. E-mail: cristinagamba@med.ucm.es

Received 9 August 2010; accepted 19 October 2010

DOI 10.1002/ajpa.21451

Published online 17 December 2010 in Wiley Online Library (wileyonlinelibrary.com).

TABLE 1. Description of analyzed skeletons^a

Tomb	Inscription	Individual	Preservation	Anthropological age estimation	Anthropological sex estimation	Sample	Type of sample
A	Martín López de Córdoba-Hinestrosa	1SC	Mummy	Infant (6–7 months)	Unknown	1SC1	Cranial fragment
						1SC2	Cranial fragment
		2SC	Mummy	Infant (1–2 months)	Unknown	2SC1	Cranial fragment
						2SC2	Cranial fragment
		3SC	Mummy	Mature/senile (55–70 years)	Male	3SC1	Tooth
B	Isabel de Zuazo					3SC2	Tooth
		4SC	Mummy	Senile (60–75 years)	Female	4SC1	Tooth
C	Urraca García de Tapia (Gabriel López de Córdoba-Hinestrosa) ^b					4SC2	Tooth
		5SC	Skeleton	Adult (35–50 years)	Male	5SC1	Tooth
						5SC2	Tooth
		6SC	Skeleton	Adult (35–50 years)	Male	6SC1	Tooth
						6SC2	Tooth
D	Alfonso García de León					7SC1	Tooth
		7SC	Mummy	Adult (30–40 years)	Male	7SC2	Tooth
						7SC3	Rib

^a Samples appear according to the tomb in which they were found. Details about the inscription and further anthropological analysis of the human remains are also given.

^b Unclear burial inscription.

1994; Caramelli et al., 2007; Dissing et al., 2007; Bogdanowicz et al., 2009; Coble et al., 2009).

In this regard, the aim of this study was to test several archaeological hypotheses combining ancient DNA kinship analysis, anthropological and archaeological data, and genealogical information. Studied samples date from the 15th Century and come from a set of tombs found at a church in Central Spain. The dry environment of the burial site favored the natural mummification of most of them and presumably protected human tissues against molecular degradation allowing nuclear marker analysis.

MATERIALS AND METHODS

Archaeological hypotheses

Samples came from San Esteban church, located in Cuellar (Segovia, Central Spain), and dated in the 15th Century. Archaeological intervention took place in 2008, with the aim of restoring four wall niches set in pairs at both sides of the apse. Five mummified individuals, three adults and two infants, and two complete adult skeletons were retrieved from the niches. Specimens were covered with a thick layer of lime, a measure normally taken in the Middle Ages in case of death from infectious disease (Palomino Lázaro et al., 2009).

Burial inscriptions found on the covers of the niches refer to the Córdoba-Hinestrosa family, which represents an important Spanish lineage, probably related to King Alfonso IX of León. A genealogical tree is provided in the supplementary material (Fig. S1, Supporting Information). A distribution of individuals in the niches is given in Table 1, together with their state of preservation, burial inscriptions, and samples studied.

One of the burials, that of Isabel de Zuazo according to the inscription (Burial B in Table 1), is particularly interesting because some remarkably well-preserved papal bulls were found together with the mummy. These refer mainly to the charity work of the deceased.

An additional problem with this burial set concerns the inscriptions. Each niche had only one intelligible inscription. However, Niches A and C contained more than one individual (Table 1). In the case of Niche C, it was possible to detect a second unclear inscription.

Archaeologists thought that this additional inscription referred to another member of the same family, Gabriel López de Córdoba-Hinestrosa, the father of Martín López de Córdoba-Hinestrosa, buried in Niche A. Moreover, Niche C had an inscription referring to a female, but included two adults anthropologically determined as males. In this case, the archaeological interest focuses also in knowing whether the genetic determination of sex coincides with the anthropological findings.

Thus, archaeological hypotheses were mainly focused on putative relationships between individuals buried together and on verifying the genealogical information contained in the inscriptions. To solve this twofold problem, our approach consisted in (1) validating anthropological sex determination with genetic data, (2) verifying if it were possible to establish a paternity relationship between individuals buried together (Burials A and C), especially in the case of two infants buried with a male adult (Burial A), (3) testing the probability of sibship between the two infants from Burial A and their possible maternity or paternity relationship with all the adult individuals, and (4) clarifying the identity of individuals buried together in Niche C, according to the inscriptions.

Anthropological analysis

After archaeological intervention, individuals were sent to the Department of Anatomy in the Faculty of Medicine of the University of Valladolid for anthropological analysis. In the case of mummified individuals, X-ray computed tomography was carried out, with the aim of obtaining three-dimensional information of skeletons for anthropological analysis (sex, height, age, and paleopathology determinations).

The age of immature specimens was determined by the degree of development in tooth bud and tooth eruption (Schour and Massler, 1941), length of long bones (Hoppa and Sanders, 1994), and chronology of bone synostosis (Brothwell, 1993).

Adult age ranges were estimated on the basis of dental attrition and degree of general ossification of the skeleton, mainly cartilage, and patterns of articular surface attrition (Campillo, 2001). Sex was determined according to Ubelaker (1978), based on classical features such

as cranial and pelvic morphology and long bones robustness.

Ancient DNA analysis

Sample selection. Sample selection was performed by E.F. and E.A. in the Department of Anatomy (Faculty of Medicine, University of Valladolid). As recommended for ancient DNA analyses, at least two samples from each skeleton were chosen.

In adult skeletons, entire teeth samples, without external fissures or caries, were selected. In immature individuals, compact bones were preferred. Fifteen samples were selected according to good macroscopic preservation. Further information about the samples is provided in Table 1. Pictures of all selected samples are available in the supplementary material.

Criteria of authenticity. Genetic analyses were carried out in specialized ancient DNA laboratories. Pre-PCR, PCR, and post-PCR procedures were carried out in three physically isolated areas located in the same building but separated by a long distance. Sample cleaning and grinding processes and DNA extraction were performed in different rooms of the same laboratory. Post-PCR sequencing analyses were carried out in a different building (Faculty of Biology, Complutense University of Madrid).

Ancient DNA laboratories were equipped with UV light lamps. Workbenches and laboratory equipment were routinely cleaned with bleach and UV-irradiated before and after each experiment. Access to these laboratories was limited to three people. In this case, all experimental analyses were conducted by a single researcher (C.G.) to reduce staff DNA contamination. All sample preparation, extraction, and PCR procedures were performed wearing disposable laboratory coveralls, masks, caps, glasses, shoe covers, and gloves. DNase- and RNase-free reagents and consumables were used.

To prevent cross-contamination, samples from the same individual were processed in separate rounds of cleaning, grinding, DNA extraction, and amplification. Possible contamination was monitored with extraction blanks, and at least three PCR negative controls were included with each of the seven samples.

Results reproducibility was assessed by setting up independent extractions and amplifications from each skeleton. DNA was extracted from at least two different samples per individual. Two independent PCR amplifications of nuclear short tandem repeats (STRs) were performed from each extract. To monitor exogenous DNA contamination of the samples, genetic profiles were recovered from all the people involved in sample manipulation, including laboratory and anthropological staff.

Sample cleaning and grinding. Sample cleaning was carried out using a Sand Blaster (Dentalarm Base 1 Plus). This equipment allows the removal of about a millimeter of the bone/tooth surface using aluminum oxide powder under pressure. The aim of this procedure is to clean the sample and remove contaminant DNA molecules from its outer surface. Samples were then irradiated with UV light for about 30 min in a laminar flow cabinet and transferred to sterile grinding vials. Grinding was performed in a Freezer Mill (SPEX Model 6700) filled with liquid nitrogen. The resulting powder was stored at -20°C until DNA extraction was performed.

DNA extraction. The entire DNA extraction process was carried out in a laminar flow cabinet in the DNA extraction laboratory. DNA was extracted from approximately 500 mg of sample powder using a modification of the protocol of Rohland and Hofreiter (2007a). In this protocol, DNA is absorbed to silica in the presence of high concentrations of a chaotropic salt. In this work, guanidinium thiocyanate (GuSCN) was replaced by sodium chloride (NaCl) at the same concentration. GuSCN was replaced by NaCl not only because it is much cheaper, but also because DNA retrieval can be higher. Although NaCl is worse in eliminating PCR-inhibitors (Rohland and Hofreiter, 2007b), we tested this method in our laboratory obtaining highly efficient results in archaeological samples (data not shown).

Nuclear DNA amplification with the AmpFISTR[®] MiniFiler kit. Eight polymorphic autosomal STRs (D13S317, D7S820, D2S1338, D21S11, D16S539, D18S51, CSF1PO, and FGA) together with the Amelogenin locus (AMEL) were PCR-amplified in the 15 DNA extracts obtained from the seven individuals of study using the AmpFISTR[®] MiniFiler kit (Applied Biosystems). All these markers are included in the Combined DNA Integrated Systems (CODIS), the international reference for forensic science markers. The AmpFISTR[®] MiniFiler kit has been specifically designed for degraded DNA and is characterized by high sensitivity, compared with classical forensic kits, and the reduction of amplicon lengths (Mulero et al., 2008). As already mentioned, two PCRs were set up for each extract for further comparison.

Amplifications were carried out on a Multigene I thermocycler (Labnet), according to the manufacturer's recommendations, with the exception of the concentration of DNA. In this case, 5 μl of DNA extract was added instead of 10 μl , to reduce possible inhibitory effects. Amplicon sizes and PCR cycling parameters are described in the supplementary material (Fig. S2 and Table S2, Supporting Information).

Capillary electrophoresis. STR amplicons were separated under standard conditions on an ABI PRISM[®] 3730 Genetic Analyzer (Applied Biosystems), located in a different building (Faculty of Biology, Complutense University of Madrid). Results were analyzed using GeneMapper[®] software, version 4.0.

Kinship analysis. To evaluate kinship between individuals, the likelihood ratio (LR) and paternity, maternity, and sibship probabilities were calculated using Familias software, version 1.7 (Egeland et al., 2000). LR is a standard statistic in forensic genetics (Jobling and Hurler, 2004), which represents a quotient between the probabilities of a same event under two different hypotheses (H_0 and H_1). The paternity index (PI) is then a special case of LR, where H_0 represents the hypothesis of coincidence between the profile of an alleged father and that of the alleged son because of a true kinship, and H_1 to a random coincidence. Values of PI can be converted into probability paternity values (W), calculated as $W = X/(X + Y)$ in which the X and Y represented probabilities of the hypotheses paternity (H_0) and nonpaternity (H_1), respectively (Gjertson et al., 2007). Similar values for maternity and sibship probabilities were also calculated with Familias.

Databases. To calculate LRs, allelic frequencies of the original population of the samples are needed. Because

TABLE 2. STRs AmpFLSTR[®] MiniFiler[™] kit results and consensus profiles

	Skeleton	Sample	PCR	D13S317	D7S820	AMEL	D2S1338	D21S11	D16S539	D18S51	CSF1PO	FGA	
Tomb A	1SC	1SC1	1	12/13	— ^a	—	18/19	—	—	—	10/11	—	
			2	—	—	—	—	—	11 ^b	—			
		1SC2	1	13 ^b	—	X ^b	18/19	—	—	—	10/11	—	
			2	12 ^b	—	X/Y	18/19	—	—	—	10/11	—	
		Consensus profile			12/13	—	X/Y	—	—	—	—	—	
	2SC	2SC1	1	11/12	—	X/Y	19 ^b	—	—	—	10/11	—	
			2	11/12	—	X/Y	17/19	—	—	—	10/11	—	
		2SC2	1	11/12	—	—	—	—	—	—	10/11	—	
			2	—	—	—	—	—	—	—	10/11	—	
		Consensus profile			11/12	—	X/Y	—	—	—	—	—	
	3SC	3SC1	1	10/12	8/9	X/Y	19/24	28/32.2	12/13	11/18	10/11	23/24	
			2	10/12	8/9	X/Y	19/24	28/32.2	12/13	11/18	10/11	23/24	
		3SC2	1	10/12	8/9	X/Y	19/24	28/32.2	12/13	11/18	10/11	23/24	
			2	10/12	8/9	X/Y	19/24	28/32.2	12/13	11/18	10/11	23/24	
		Consensus profile			10/12	8/9	X/Y	19/24	28/32.2	12/13	11/18	10/11	23/24
Tomb B	4SC	4SC1	1	11	8	X	24 ^b	—	—	—	11	22	
			2	11	—	X	—	28/30	—	15/16	11	22	
		4SC2	1	11	—	X	20 ^b	—	—	13/16	11	—	
			2	11	—	X	20/24	—	—	13/16	11	22	
		Consensus profile			11/11	8/8	X/X	20/24	28/30	—	13/16	11/11	22/22
Tomb C	5SC	5SC1	1	9/11	10/12	X	17/25	28/30	13	14/17	10/12	22/23	
			2	9/11	10/12	X	17/25	28/30	13	14/17	10/12	22/23	
		5SC2	1	9/11	10/12	X	17/25	28/30	13	14/17	10/12	22/23	
			2	9/11	10/12	X	17/25	28/30	13	14/17	10/12	22/23	
		Consensus profile			9/11	10/12	X/X	17/25	28/30	13/13	14/17	10/12	22/23
	6SC	6SC1	1	10/11	8/10	X/Y	19/24	28/32.2	9/12	17/18	11/12	24	
			2	10/11	8/10	X/Y	19/24	28/32.2	9/12	17/18	11/12	24	
		6SC2	1	10/11	8/10	X/Y	19/24	28/32.2	9 ^b	17/18	11/12	24	
			2	10/11	8/10	X/Y	19/24	28/32.2	9 ^b	17/18	11/12	24	
Consensus profile			10/11	8/10	X/Y	19/24	28/32.2	9/12	17/18	11/12	24/24		
Tomb D	7SC	7SC1	1	11/13	8/12	X/Y	24/25	29/32	9	12/15	11/12	19/23	
			2	11/13	8/12	X/Y	24/25	29/32	9	12/15	11/12	19/23	
		7SC2	1	11/13	8/12	X/Y	24/25	29/32	9	12/15	11/12	19/23	
			2	11/13	8/12	X/Y	24/25	29/32	9	12/15	11/12	19/23	
		7SC3	1	11/13	—	X/Y	24/25	29/32	—	12/15	12 ^b	19 ^b	
			2	—	—	—	—	—	—	—	12 ^b	—	
		Consensus profile			11/13	8/12	X/Y	24/25	29/32	9/9	12/15	11/12	19/23

^a Dashes indicate lack of results.^b Indicate allelic dropout.

of the absence of other ancient DNA data corresponding to the same period and region, a modern Spanish DNA population database was used. This database was compiled by the Laboratory of Forensic and Population Genetics (Complutense University of Madrid) for routine forensic casework. Possible differences between ancient allelic frequencies and modern ones were taken into account in the interpretation of the results.

RESULTS

Authenticity of the results

We did not find any exogenous contamination in extraction and PCR blanks, supporting the authenticity of results. In addition, the absence of mixed STR profiles suggested that no cross-contamination had occurred.

Moreover, the same loci analyzed for the ancient specimens were also typed for all the staff who manipulated the samples: archaeologists, anthropologists and laboratory staff. Results for modern samples are provided in supplementary material (Table S1, Supporting Information). Ancient profiles were compared with modern, and modern DNA contamination may be rejected.

STR profiles

At least two extracts for each specimen were obtained. A consensus genotype was reconstructed taking into account the common alleles obtained from two PCRs per extract. The genotypes obtained for each sample and amplification are reported in Table 2.

At some amplifications from the same sample and marker, marked with superscript "b" in Table 2, only one of the two alleles identified in a previous amplification was detected. This was interpreted as an allelic dropout phenomenon, consisting of the random amplification of one allele in a heterozygous sample (Butler, 2005). This event is frequent in degraded samples, with low DNA content.

Complete or partial profiles were obtained for all the individuals. The five mature specimens provided results for all the nine STRs. Skeleton 4SC, whose preservation was worse than that of the other specimens, yielded only eight markers. Moreover, two STRs of 4SC did not yield replicable results (D7S820 and D21S11 markers), and, for D7S820, we obtained only one homozygous amplification. Even if this individual is considered homozygous for D7S820 in the consensus profile, for further analysis, we took into account the possibility of the phenomenon of allelic dropout.

TABLE 3. Molecular sex determination of studied samples and their tomb distribution. Anthropological age and sex estimations are also included for comparison

Tomb	Individual	Preservation	Anthropological age	Anthropological sex	Molecular sex
A	1SC	Mummy	Infant (6–7 months)	Unknown	Male
	2SC	Mummy	Infant (1–2 months)	Unknown	Male
	3SC	Mummy	Mature/senile (55–70 years)	Male	Male
B	4SC	Mummy	Senile (60–75 years)	Female	Female
C	5SC	Skeleton	Adult (35–50 years)	Male	Female
	6SC	Skeleton	Adult (35–50 years)	Male	Male
D	7SC	Mummy	Adult (30–40 years)	Male	Male

Concerning the two immature specimens (skeletons 1SC and 2SC), only five of nine genetic systems were typed, and we detected low replication rate and allelic dropout in several markers (Table 2).

Sex determination

As can be seen in Table 2, amplification of the AMEL was successful in all the analyzed individuals. Molecular and anthropological sex determinations are compared in Table 3. Genetic data matched anthropological determinations in all adult skeletons except for 5SC, which was anthropologically identified as male adult. However, 5SC was genetically typed as a female, in accordance with the niche inscription. Anthropological sex determination of the two infants was not clear because of their very young age (only few months). Amelogenin amplification was successful in both cases, revealing their male sex.

Kinship relationships

We tested archaeological hypotheses comparing the obtained STRs profiles. We first investigated possible relationships between individuals buried together. For Niche A, we compared the male mummy (3SC) profile with the two infants buried together (1SC and 2SC). Because of partial profiles obtained for the last two, we compared only five of nine markers. Paternity was excluded because of the extreme differences found for the alleles in D18S51 system. The probability of sibship between the two infants was calculated and yielded non-conclusive values of 2.11% and 68% for LR and W, respectively. Also, we investigated paternity and maternity relationships among infant individuals and adults found in other niches. The female adult individual 5SC showed a LR value of 2.11 with the infant skeleton 2SC, corresponding to a probability of maternity (W) of about 68%. It represents a low kinship value, mainly due to the incomplete profiles obtained for both immature specimens.

In Niche C, a possible paternity relationship on the bases of archaeological hypotheses was tested. In this case, there were three inconsistencies in D2S1338, D16S539, and FGA loci. Then we compared all adult profiles among them to detect possible paternal relationships, mainly focusing on possible relationships suggested by genealogical and archaeological data such as unclear niche inscriptions. Samples 3SC and 6SC, from which complete genotypes were recovered, yielded a relatively high value of PI, 4,422.24, which corresponds to $W = 99.98\%$. These values are close to those obtained in routine forensic casework, and they may indicate a paternity relationship between 3SC and 6SC. Other comparisons are not cited because there was at least one inconsistency between loci.

DISCUSSION

Efficiency in recovering nuclear DNA profiles

The main aim of this work has been achieved, because the high efficiency in recovering DNA from ancient materials allowed us to integrate anthropological, archaeological, and genealogical data together with the molecular kinship analysis. We obtained genotypes for seven ancient specimens and, subsequently, three different degrees of kinship relationship could be determined between them.

First, almost complete typing of samples indicated exceptional ancient nuDNA preservation. Thus, we can assume that highly favorable environmental conditions for DNA preservation were present in burials (Poinar et al., 2003). Specifically, the lime covering the bodies could have created a dry environment, leading to natural mummification. Mummified tissues have been proved to yield interesting ancient DNA results. This is the case of the ancient Egyptian mummies from the burial cluster of Tutankhamun's family, which were successfully typed (Hawass et al., 2010).

On the other hand, the results obtained suggest that the combination of a silica-based DNA extraction protocol together with the use of specific amplification kits for degraded DNA is the best strategy for the analysis of nuclear markers. Both the modification introduced in the extraction protocol, i.e., the use of NaCl instead of GuSCN in the DNA binding step to the silica, and the specific formulation of the AmpFISTR[®] MiniFiler kit could have increased DNA recovery and overcome an inhibitory effect present in some of the samples.

Similar strategies, such as the combined silica-based extraction method and AmpFISTR[®] MiniFiler kit, gave successful nuDNA results in 7th Century human remains from Bavaria, Germany (Vanek et al., 2009). Autosomal DNA amplification strategies, before the AmpFISTR[®] MiniFiler Kit, like forensic kits AmpFISTR[®] ProfilerPlus[®] or Identifiler[®] (Applied Biosystems), were less efficient because of the length of amplicons (Ricaud et al., 2005b).

Sex determination

With just one exception, anthropological data were consistent with molecular sex determination. Skeleton number 5SC was anthropologically classified as male, but showed only an X chromosome that probably belonged to Urraca García de Tapia (a female name), as shown in the inscription of the niche. Thus, this result leads to correcting the sex assignment based on anthropological data (two males in Niche C) and is consistent with the correspondent historical assumption referring to the inscriptions. The issue of the extra body within the Urraca niche is discussed in the next paragraph, concerning kinship analysis. Moreover, it was possible to

assign a sex to the two immature skeletons, not anthropologically classified.

Kinship relationships

Genetic analyses enabled establishing relationships between individuals on the basis of archaeological hypotheses. Archaeologists suggested that individual 3SC was the father of the two infants (1SC and 2SC) buried with him. Genetic results do not support this paternity relationship. Archaeologists also wonder whether these two immature mummies could be two brothers related with some other adult of the niche set. We determine a 68% probability of sibship between them, but it was also possible to relate the adult female 5SC to one of them (2SC), with a probability of maternity of 68%. We separately investigated and rejected 5SC maternity relationship with 1SC because of the presence of two exclusions of four comparable autosomal markers. Nevertheless, these percentage values are clearly low according to routine forensic standards; each one can represent a clue supporting different archaeological hypotheses: (1) 1SC and 2SC could belong from different parents and 5SC could be the mother of 2SC with a 68% probability; (2) 1SC and 2SC could be siblings, but their parents are not buried in the studied niche set.

Genetic results also disproved paternity relationship between the adult individuals 5SC and 6SC, buried together in Niche C. On the other hand, the unclear inscription on Niche C suggested the presence of Gabriel López de Córdoba-Hinestrosa within, who is the father of Martín López de Córdoba-Hinestrosa, the individual 3SC, buried in Niche A. Genealogical and archaeological data, therefore, suggested a possible paternity kinship between 3SC and the male individual buried in Niche C (6SC). This paternity relationship hypothesis between 3SC and 6SC has been strongly confirmed by STRs genotyping (4,422.24 and 99.98% for LR and W, respectively).

CONCLUSIONS

Our results point at the importance of ancient DNA contribution in archaeological kinship studies. Genetic profiles shed some light on the purely hypothetical assumptions of the archaeologists, mainly in the case where the paternity hypothesis was supported and in those cases where close genetic relationships were excluded. The results also indicate the usefulness of molecular sex determination in confirming anthropological findings and providing new evidence in unsuccessful anthropological analyses, such as the ones performed on immature skeletons. In general, we conclude that technical advances in forensic genetics, such as the AmpFISTR® MiniFiler kit, represent a useful tool in ancient DNA studies. However, only a multidisciplinary approach, where forensic genetics technology, classical anthropology, and archaeological research work are integrated, can yield valuable results for the understanding of burial patterns and population social structure in historical events.

ACKNOWLEDGMENTS

The authors thank Ángel Palomino Lázaro and Manuel Moratinos García, from Aratikos Arqueólogos society, for providing samples and archaeological information. They also thank all anthropologists involved in

this study, Mercedes Barbosa Cachorro, Félix de Paz Fernández, María García Velasco, and Eva Ferrero Infesta.

LITERATURE CITED

- Amory S, Keyser C, Crubézy E, Ludes B. 2007. STR typing of ancient DNA extracted from hair shafts of Siberian mummies. *Forensic Sci Int* 166:218–229.
- Anderung C, Persson P, Bouwman A, Elburg R, Götherström A. 2008. Fishing for ancient DNA. *Forensic Sci Int Genet* 2:104–107.
- Bogdanowicz W, Allen M, Branicki W, Lembring M, Gajewska M, Kupiec T. 2009. Genetic identification of putative remains of the famous astronomer Nicolaus Copernicus. *Proc Natl Acad Sci USA* 106:12279–12282.
- Brothwell D. 1993. Desenterrando huesos. La excavación, tratamiento y estudio de restos del esqueleto humano. México: Fondo de Cultura Económica.
- Butler JM. 2005. Forensic DNA typing. Biology, technology, and genetics of STR markers. Burlington: Elsevier Academic Press.
- Campillo D. 2001. Introducción a la Paleopatología. Barcelona, Spain: Bellaterra.
- Caramelli D, Lalueza-Fox C, Capelli C, Lari M, Sampietro ML, Gigli E, Milani L, Pilli E, Guimaraes S, Chiarelli B, Marin VT, Casoli A, Stanyon R, Bertranpetit J, Barbujani G. 2007. Genetic analysis of the skeletal remains attributed to Francesco Petrarca. *Forensic Sci Int* 173:36–40.
- Coble MD, Loreille OM, Wadhams MJ, Edson SM, Maynard K, Meyer CE, Niederstätter H, Berger C, Berger B, Falsetti AB, Gill P, Parson W, Finelli LN. 2009. Mystery solved: the identification of the two missing Romanov children using DNA analysis. *PLoS One* 4:e4838.
- Dissing J, Binladen J, Hansen A, Sejrson B, Willerslev E, Lynnerup N. 2007. The last Viking King: a royal maternity case solved by ancient DNA analysis. *Forensic Sci Int* 166:21–27.
- Egeland T, Mostad PF, Mevag B, Stenersen M. 2000. Beyond traditional paternity and identification cases. Selecting the most probable pedigree. *Forensic Sci Int* 110:47–59.
- Gill P, Ivanov PL, Kimpton C, Piercy R, Benson N, Tully G, Evett I, Hagelberg E, Sullivan K. 1994. Identification of the remains of the Romanov family by DNA analysis. *Nat Genet* 6:130–135.
- Gjertson DW, Brenner CH, Baur MP, Carracedo A, Guidet F, Luque JA, Lessig R, Mayr WR, Pascali VI, Prinz M, Schneider PM, Morling N. 2007. ISFG: recommendations on biostatistics in paternity testing. *Forensic Sci Int Genet* 1:223–231.
- Hawass Z, Gad YZ, Ismail S, Khairat R, Fathalla D, Hasan N, Ahmed A, Elleithy H, Ball M, Gaballah F, Wasef S, Fateen M, Amer H, Gostner P, Selim A, Zink A, Pusch CM. 2010. Ancestry and pathology in King Tutankhamun's family. *JAMA* 303:638–647.
- Higuchi R, Bowman B, Freiberger M, Ryder O, Wilson A. 1984. DNA sequences from the quagga, an extinct member of the horse family. *Nature* 312:282–284.
- Hoppa RD, Sanders SR. 1994. The dl method for examining bone growth in juveniles: a reply. *Int J Osteoarchaeol* 4:261–263.
- Hummel S, Bramanti B, Schultes T, Kahle M, Haffner S, Herrmann B. 2000. Megaplex DNA typing can provide a strong indication of the authenticity of ancient DNA amplifications by clearly recognizing any possible type of modern contamination. *Anthropol Anz* 58:15–21.
- Jobling MA, Hurles M. 2004. Human evolutionary genetics: origins, peoples & disease. New York: Garland Science.
- Keyser-Tracqui C, Crubézy E, Ludes B. 2003. Nuclear and mitochondrial DNA analysis of a 2,000-year-old necropolis in the Egyin Gol Valley of Mongolia. *Am J Hum Genet* 73:247–260.
- Loreille O, Diegoli T, Irwin J, Coble M, Parsons T. 2007. High efficiency DNA extraction from bone by total demineralization. *Forensic Sci Int Genet* 1:191–195.

- Mulero JJ, Chang CW, Lagace RE, Wang DY, Bas JL, McMahon TP, Hennessy LK. 2008. Development and validation of the AmpFISTR MiniFiler PCR amplification kit: a MiniSTR multiplex for the analysis of degraded and/or PCR inhibited DNA. *J Forensic Sci* 53:838–852.
- Pääbo S. 1985. Molecular cloning of ancient Egyptian mummy DNA. *Nature* 314:644–645.
- Pääbo S. 1986. Molecular genetic investigations of ancient human remains. *Cold Spring Harb Symp Quant Biol* 51(Pt 1):441–446.
- Pääbo S, Poinar H, Serre D, Jaenicke-Despres V, Hebler J, Rohland N, Kuch M, Krause J, Vigilant L, Hofreiter M. 2004. Genetic analyses from ancient DNA. *Annu Rev Genet* 38:645–679.
- Palomino Lázaro AL, Moratino García M, Pastor Vázquez JF. 2009. Retazos de ayer, Vol. 37. Patrimonio: Fundación del patrimonio histórico de Castilla y León. p51–58.
- Poinar H, Kuch M, McDonald G, Martin P, Pääbo S. 2003. Nuclear gene sequences from a late Pleistocene sloth coprolite. *Curr Biol* 13:1150–1152.
- Rasmussen M, Li Y, Lindgreen S, Pedersen JS, Albrechtsen A, Moltke I, Metspalu M, Metspalu E, Kivisild T, Gupta R, Bertalan M, Nielsen K, Gilbert MT, Wang Y, Raghavan M, Campos PF, Kamp HM, Wilson AS, Gledhill A, Tridico S, Bunce M, Lorenzen ED, Binladen J, Guo X, Zhao J, Zhang X, Zhang H, Li Z, Chen M, Orlando L, Kristiansen K, Bak M, Tommerup N, Bendixen C, Pierre TL, Grønnow B, Meldgaard M, Andreasen C, Fedorova SA, Osipova LP, Higham TF, Ramsey CB, Hansen TV, Nielsen FC, Crawford MH, Brunak S, Sicheritz-Pontén T, Villems R, Nielsen R, Krogh A, Wang J, Willerslev E. 2010. Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* 463:757–762.
- Ricaud F, Fedoseeva A, Keyser-Tracqui C, Crubézy E, Ludes B. 2005a. Ancient DNA analysis of human Neolithic remains found in northeastern Siberia. *Am J Phys Anthropol* 126:458–462.
- Ricaud F, Keyser-Tracqui C, Crubézy E, Ludes B. 2005b. STR-genotyping from human medieval tooth and bone samples. *Forensic Sci Int* 151:31–35.
- Rohland N, Hofreiter M. 2007a. Ancient DNA extraction from bones and teeth. *Nat Protoc* 2:1756–1762.
- Rohland N, Hofreiter M. 2007b. Comparison and optimization of ancient DNA extraction. *Biotechniques* 42:343–352.
- Schour I, Massler M. 1941. The development of the human dentition. *J Am Dent Assoc* 28:1153–1160.
- Schultes T, Hummel S, Herrmann B. 2000. Ancient DNA-typing approaches for the determination of kinship in a disturbed collective burial site. *Anthropol Anz* 58:37–44.
- Sprecher CJ, McLaren RS, Rabbach D, Krenke B, Ensenberger MG, Fulmer PM, Downey L, McCombs E, Storts DR. 2009. PowerPlex[®] ESX and ESI Systems: a suite of new STR systems designed to meet the changing needs of the DNA-typing community. *Forensic Sci Int Genet* 2:2–4.
- Ubelaker DH. 1978. Human skeletal remains: excavation, analysis, interpretation. Chicago, IL: Aldine Publishing Co.
- Vanek D, Saskova L, Koch H. 2009. Kinship and Y-chromosome analysis of 7th century human remains: novel DNA extraction and typing procedure for ancient material. *Croat Med J* 50:286–295.
- Wiesner R, Rüegg J, Morano I. 1992. Counting target molecules by exponential polymerase chain reaction: copy number of mitochondrial DNA in rat tissues. *Biochem Biophys Res Commun* 183:553–559.

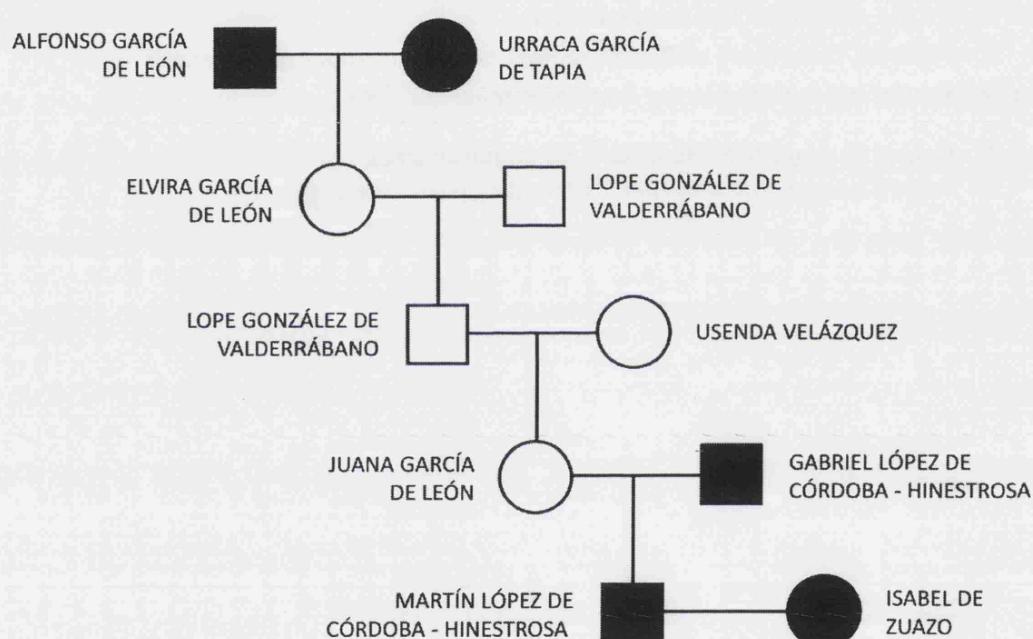
SUPPLEMENTARY MATERIAL

The supplementary material includes:

1. **Figure S1. Genealogical tree of the Córdoba-Hinestrosa family**
2. **Samples pictures per individual**
 - a. Individual 1SC, samples 1SC1 and 1SC2
 - b. Individual 2SC, samples 2SC1 and 2SC2
 - c. Individual 3SC, samples 3SC1 and 3SC2
 - d. Individual 4SC, samples 4SC1 and 4SC2
 - e. Individual 5SC, samples 5SC1 and 5SC2
 - f. Individual 6SC, samples 6SC1 and 6SC2
 - g. Individual 7SC, samples 7SC1 and 7SC2
 - h. Individual 7SC, samples 7SC3
3. **Table S1. Laboratory staff, archaeologists and anthropologists genetic profiles for the loci included in the AmpFSTR®MiniFiler™ Kit**
4. **Figure S2 AmpFSTR®MiniFiler™ Kit amplicon sizes**
5. **Table S2 AmpFSTR®MiniFiler™ Kit cycling parameters**

FIGURE S1

Genealogical tree of the Córdoba-Hinestrosa family. Figures in black belong to individuals with burial inscriptions. Inscription of Gabriel López de Córdoba-Hinestrosa was unclear (Male = square, Female = circle). This tree has been drawn on the basis of information provided in DE CEBALLOS-ESCALERA Y GILA, A. 1986. Los sepulcros gótico-mudéjares de San Esteban de Cuéllar. Boletín de la Sociedad Segoviana de Heráldica y Genealogía "don Gaspar Ybáñez de Segovia", 1: 21.



SAMPLES PICTURES



TABLE S1

Genetic profiles of the laboratory staff, archaeologists and anthropologists for the loci included in the AmpFISTR® MiniFiler™ Kit.

Staff	Code	D13S317	D7S820	AMEL	D2S1338	D21S11	D16S539	D18S51	CSF1P0	FGA
Lab staff	CG	9/11	8/11	X/X	17/24	31/32.2	11/12	12/12	12/12	21/22
	MT	8/11	9/11	X/X	23/23	30/31.2	9/11	12/12	10/10	24/25
	EF	11/13	10/10	X/X	24/24	28/28	10/13	14/19	12/12	24/24
	EA	8/12	8/10	X/Y	20/23	30/31.2	13/13	12/16	11/12	20/21
Archaeologists/ Anthropologists	FPV	12/12	8/12	X/Y	16/17	30/30	11/12	13/17	11/12	19/21
	FPF	9/11	9/9	X/Y	21/21	28/31.2	11/13	15/18	10/10	20/22
	MGV	11/12	8/8	X/X	20/24	24.2/32	11/12	11/17	11/12	19/22
	EFI	11/11	6/9	X/X	16/17	29/31.2	10/13	12/4	10/11	21/22
	MBC	12/12	11/12	X/X	18/21	29/30	12/12	12/12	10/12	20/21

FIGURE S2.

AmpFSTR®MiniFiler™ Kit amplicon sizes. This Figure has been drawn on the basis of information provided in the *AmpFSTR®MiniFiler™* Kit user guide.

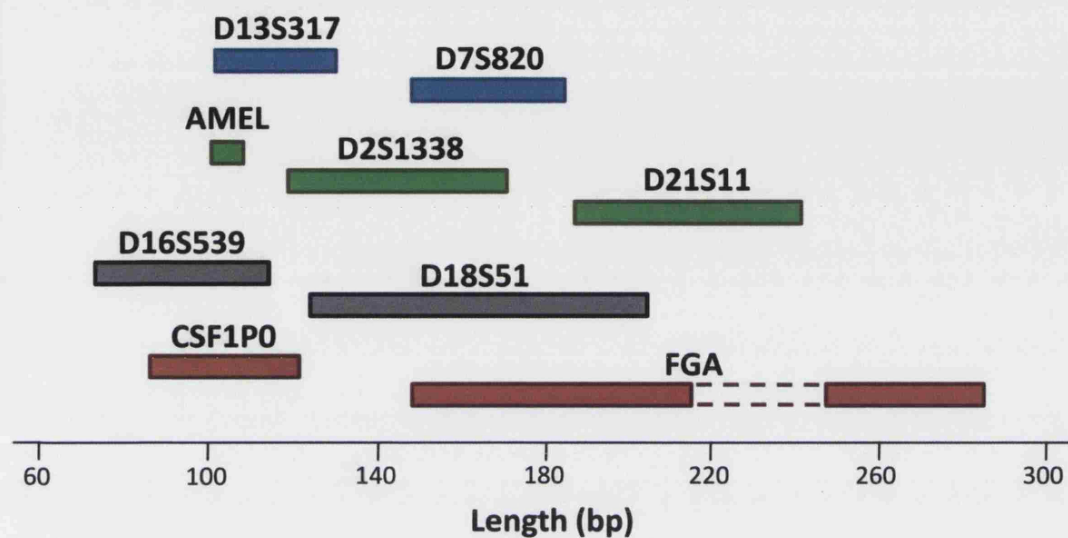


TABLE S2

AmpFSTR®MiniFiler™ Kit PCR cycling. Amplifications were carried out according to the manufacturer's recommendations (see *AmpFSTR®MiniFiler™* Kit user guide).

PCR Steps		Temperature	Time
Initial Incubation		95 °C	11 min
30 cycles	Denature	94 °C	20 sec
	Anneal	59 °C	2 min
	Extend	72 °C	1 min
Final Extension		60 °C	45 min
Hold		4 °C	∞

DISCUSIÓN

DISCUSIÓN

CONSIDERACIONES GENERALES

El objetivo principal de esta tesis doctoral es el análisis y validación de resultados en el campo del ADN antiguo, precedido de la adecuación y mejora de las técnicas disponibles. Como ya se ha comentado en el capítulo de **Objetivos**, este trabajo se ha desarrollado a lo largo de tres ejes principales: (1) desarrollo metodológico, (2) estudio diacrónico-poblacional y (3) establecimiento de relaciones familiares.

En primer lugar, se evaluaron e implementaron en el laboratorio diferentes técnicas para el análisis de muestras antiguas y en paralelo se estimó el impacto de diferentes factores que pueden influir en la obtención de resultados, con un fin prospectivo (**ARTÍCULOS 1 y 2**). A continuación se aplicaron diferentes técnicas experimentales y de análisis de datos en función de la problemática concreta. Las dos aplicaciones que se presentan en esta tesis incluyen el estudio diacrónico de los diferentes pobladores del noreste de la Península Ibérica (**ARTÍCULOS 3 y 4**) y la determinación del sexo y de relaciones familiares en un conjunto de individuos enterrados en la Iglesia de San Esteban de Cuéllar, Segovia (**ARTÍCULO 5**).

En este capítulo se discuten y analizan punto por punto los objetivos planteados en la presente tesis doctoral (ver capítulo de **Objetivos**). Dentro de cada uno de los objetivos planteados se desarrollan diferentes apartados, especificando el artículo al que se hace referencia con la correspondiente numeración. Para facilitar la lectura de este capítulo, algunas de las imágenes y tablas contenidas en los artículos se vuelven a representar a lo largo de la discusión de los resultados.

DISCUSIÓN DE LOS OBJETIVOS – METODOLOGÍA Y VALIDACIÓN DE LOS RESULTADOS

D.1 Comparación de diferentes metodologías disponibles en la literatura de ADN antiguo y forense para el aislamiento y la amplificación de ADN de muestras arqueológicas. Selección de aquellos protocolos más eficientes y que se ajustan a las características de las muestras disponibles y a las finalidades concretas del estudio.

El desarrollo de esta tesis doctoral se llevó a cabo en un grupo de investigación con una experiencia de más de diez años en el campo del ADN antiguo (aDNA) (Arroyo-Pardo et al., 2002; García-Bour et al., 2004; Fernández, 2005; Fernández et al., 2009). Este bagaje de experiencia permitió el comienzo del trabajo de tesis en un laboratorio con infraestructuras y protocolos adecuados para el análisis de muestras antiguas.

El trabajo se ha desarrollado teniendo en cuenta que el aDNA es una disciplina en continua evolución gracias al desarrollo de nuevas técnicas específicamente diseñadas para muestras degradadas. También contribuye a su desarrollo la implementación de metodologías diseñadas para disciplinas afines, como la genética forense. A lo largo del desarrollo inicial de la tesis doctoral, se introdujeron algunas de las medidas y protocolos propuestos en la literatura reciente para mejorar la obtención de resultados a partir de material antiguo (ver **ARTÍCULO 1**). Además, la posibilidad de realizar una estancia de tres meses en otro laboratorio especializado en ADN antiguo (*Laboratoire d'Anthropologie des Populations du Passé (LAPP) – De la Préhistoire à l'Actuel: Culture, Environnement et Anthropologie (PACEA), Université de Bordeaux 1 – CNRS, Talence, Francia*) permitió comparar protocolos e infraestructuras. Así se pudieron introducir algunas modificaciones en los protocolos del laboratorio de origen y sugerir otras en el laboratorio de acogida. A continuación se destaca la importancia de la metodología y se describen las modificaciones introducidas en las diferentes fases del análisis.

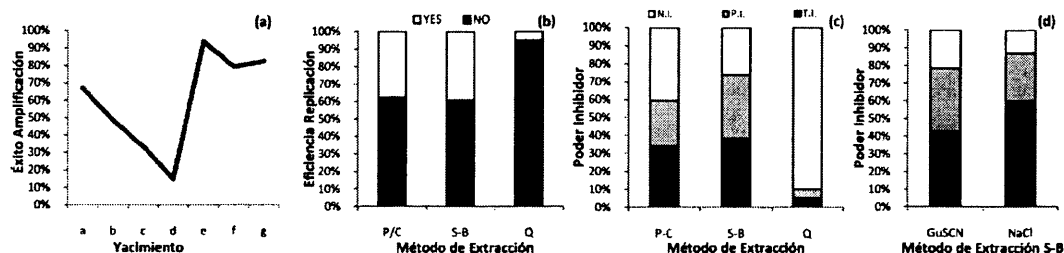
D.1.1 COMPARACIÓN DE DIFERENTES METODOLOGÍAS PARA LA EXTRACCIÓN DE ADN (ARTÍCULO 1)

La extracción de ADN es una de las fases más delicadas y resulta muy sensible a posibles contaminaciones exógenas procedentes del manipulador, del ambiente de trabajo, de los reactivos o de otras muestras analizadas con anterioridad (contaminación por arrastre o *carry-over*) (ver **Apartado I.3.3.3**). La detección de esta clase de contaminaciones es muy importante, puesto que en fases sucesivas del análisis podrían interpretarse erróneamente como resultados endógenos a las muestras. Por otra parte, actualmente no existe un protocolo común y estandarizado entre los diferentes laboratorios de aDNA. Por todo ello, en el presente trabajo se probó la eficiencia de cuatro protocolos de extracción populares en el campo del ADN antiguo y de la genética forense en un conjunto de muestras humanas antiguas de diferente procedencia.

Para evaluar la eficiencia de los diferentes protocolos de extracción ensayados, se midió la eficiencia en la obtención de resultados de cada uno de ellos, desglosándola en las distintas fases del análisis: amplificación (porcentaje de amplificaciones positivas sobre el total), secuenciación (porcentaje de secuencias legibles, no contaminadas por los manipuladores) y replicación (porcentaje de secuencias replicadas a partir de la misma muestra). Otra variable que se empleó también para estimar la eficiencia de los diferentes protocolos es la capacidad de los mismos para eliminar moléculas inhibitoras. Para ello se midió el poder inhibitor de los distintos extractos obtenidos con los diferentes métodos añadiendo 2,5 µl de extracto de ADN a una PCR estándar y observando la presencia/ausencia de amplificación en un gel de agarosa. Como se ha puesto de manifiesto en muchas publicaciones (Pääbo et al., 1989; Tuross, 1994; Wilson, 1997), la inhibición es una de las principales causas de fracaso de los estudios de DNA antiguo y por ello se ha evaluado su impacto.

Al ensayar diferentes protocolos fue posible destacar como algunos no son aptos para su implementación en ADN antiguo, mientras otros resultan eficaces (ver la **Figura D 1**).

Figura D 1 (a) Eficiencia en la amplificación ("Eficiencia PCR") frente a los diferentes yacimientos arqueológicos analizados, ordenados de la letra "a" a la "g" en orden cronológico de más antiguo a más reciente. (b) Eficiencia en la obtención de resultados replicables (media de la "Eficiencia 3,4" y "Eficiencia 5,6") frente al método de extracción. (c) Poder inhibitor de los extractos en función del método de extracción empleado (N.I.=Sin Inhibición, P.I.=Inhibición Parcial, T.I.=Inhibición Total). (d) Poder de inhibición frente a los dos métodos de extracción basados en el empleo de sílica, uno incluye GuSCN y otro NaCl. P/C: protocolo basado en el empleo de fenol/cloroformo (Fernández 2005); S-B: protocolo basado en el empleo de sílica (Rohland y Hofreiter 2009); Q: protocolo QIAamp DNA Micro Kit (Qiagen) (Figura 1 del ARTÍCULO 1).



El protocolo desarrollado por Yang y colaboradores (Yang et al., 1998) y modificado por Bouwman y Brown (2002) se incluyó debido a que fue concebido específicamente para su aplicación en el campo del aDNA (ver **Material y Métodos del ARTÍCULO 1**). Este protocolo presenta la ventaja de que, por un lado, evita el empleo de solventes orgánicos y, por otro, reduce la contaminación con ADN fresco, dado que emplea columnas con una matriz de sílica que impiden la purificación de moléculas más largas de 10 kb. Este protocolo fue seleccionado como el más eficiente y rápido por un equipo especializado en aDNA dentro de un conjunto de cinco métodos basados en el empleo de sílica (Bouwman y Brown, 2002). Además, proporcionó resultados positivos en diferentes trabajos, como en mamíferos siberianos de más de 45.000 años (Blow et al., 2008) y caballos datados en 4.000-2.000 AP (Cai et al., 2009). En el trabajo de Malmström y colaboradores (Malmström et al., 2005), se analizaron huesos neolíticos de *Canis familiaris* (3.300-2.500 AP) y también resultó eficiente, pero en este caso se detectó una elevada concentración de moléculas contaminantes.

Sin embargo, en otro estudio de muestras humanas prehistóricas (6.400-1.200 BP) de Grecia (Chilvers et al., 2008) y también en la presente tesis doctoral (ver **Resultados y Discusión del ARTÍCULO 1**), este protocolo no produjo ningún resultado. En el caso del trabajo de Chilvers y colaboradores (Chilvers et al., 2008) no se puso en discusión la validez del método y la ausencia de resultados fue atribuida a la elevada degradación

de las muestras analizadas. En el caso de la presente tesis doctoral, la ausencia de resultados para este protocolo puede atribuirse al volumen reducido en el que se desarrolla la lisis inicial, según las modificaciones introducidas por Bowman y Brown (2002). Alrededor de 500 mg de polvo de hueso o diente se digirieron en tan solo 1 ml de solución, lo que probablemente conllevó una lisis parcial y una liberación de pocas moléculas de ADN en la solución. Esto explicaría su mejor eficiencia en muestras más recientes (Cai et al., 2009; Rogaev et al., 2009) o en muestras bien preservadas (Blow et al., 2008).

El protocolo Q (*QIAamp DNA Micro Kit* de *Qiagen*) ha sido especialmente diseñado para la extracción de ADN a partir de muestras degradadas. En el campo forense ha sido ampliamente empleado sobre diferentes tipos de muestras: manchas de sangre (Xu et al., 2011), uñas (Allouche et al., 2008) y frotis bucal (Swinfield et al., 2009). Este protocolo ha sido implementado también en aDNA proporcionando resultados satisfactorios para diferentes tipos de muestras: mtDNA en pieles neolíticas de *Capra hircus* de los Alpes suizos (Schlumbaum et al., 2010) y gen de la amelogenina en muestras eneolíticas de la República Checa (Vaňharová y Drozdová, 2008). Sin embargo, existen trabajos que destacan su baja eficiencia frente a otros métodos de extracción de ADN a partir de muestras forenses (Sewell et al., 2008; Hashiyada et al., 2009).

En la presente tesis doctoral se ha podido comprobar que la mayoría de los resultados obtenidos con este protocolo se correspondían con moléculas contaminantes procedentes esencialmente del investigador (ver **Resultados y Discusión del ARTÍCULO 1**). Una explicación reside en la elevada manipulación que implica este protocolo. Un volumen de unos 10 ml (solución de lisis) se carga y centrifuga en una columna con membrana de sílica de 750 µl de forma reiterada, para hacer pasar todo el volumen. Este proceso podría provocar la saturación de la columna, impidiendo la absorción del ADN a la matriz de sílica tras la segunda o tercera elución. La baja eficiencia de este *kit* en la recuperación de moléculas endógenas antiguas también puede estar relacionada con su baja especificidad para el análisis de huesos o dientes. El protocolo proporcionado por la casa comercial presenta variaciones para los diferentes tipos de

muestras forenses y el material óseo estaría incluido en el protocolo diseñado para tejidos en general y no para huesos o dientes en particular. Además está pensado para pequeñas cantidades (se recomienda el empleo de hasta 10 mg de tejido), mientras que en el campo del aDNA se suele trabajar con alrededor de 500 mg para obtener una cantidad suficiente de ADN. Por otra parte, así como las columnas del protocolo de Yang y colaboradores (1998) impedían la purificación de moléculas de un tamaño superior a 10 kb, las incluidas en el *kit* del protocolo Q permiten la purificación de fragmentos de gran tamaño, incluso superiores a 30 kb. Todos estos factores habrían podido contribuir al aumento de moléculas contaminantes introducidas durante la extracción de ADN.

También es posible destacar la escasa presencia de moléculas inhibidoras en los extractos obtenidos con el protocolo Q (ver el panel **b** de la **Figura D 1**), algo que sucede también en el protocolo de Yang (resultados no mostrados). A priori, se podrían interpretar estos resultados como evidencia de la eficiencia de estos métodos en la eliminación de moléculas inhibidoras en muestras de ADN antiguo. Sin embargo, a la luz del fracaso de ambos métodos de extracción con nuestras muestras, la escasez de inhibidores probablemente esté indicando la baja concentración de ADN endógeno, con el que en algunos casos co-purifican este tipo de moléculas.

Se testaron también otros dos protocolos de extracción de ADN no basados en el empleo de *kits* comerciales. El primero de ellos consiste en una extracción orgánica con Fenol-Cloroformo (P/C) que se venía empleando de forma rutinaria en el laboratorio (Fernández, 2005), dado que había sido utilizado con éxito en el campo del aDNA desde sus comienzos (Pääbo y Wilson, 1988; Perry et al., 1988; Pääbo, 1990; Lee et al., 1991; Kurosaki et al., 1993; Faerman et al., 1995; Hänni et al., 1995). El segundo, publicado en *Nature Protocols* por Rohland y Hofreiter (2007 a), consiste en una extracción por absorción a una matriz de sílica preparada en laboratorio. Este protocolo fue ensayado debido a su especial éxito en la recuperación de ADN a partir de muestras muy antiguas (Rohland y Hofreiter, 2007 a). La publicación del mismo de forma detallada facilitó su implementación en el laboratorio. La validez de estos dos

últimos protocolos queda reflejada en su eficiencia para la replicación de resultados, que se sitúa en torno al 40 % (ver panel **b** de la **Figura D 1**).

Como puede verse en panel **c** de la **Figura D 1**, en el presente trabajo de tesis se ha demostrado que el protocolo P/C resulta eficiente en la eliminación de inhibidores de la PCR. Ello se debe a que la mayoría de los inhibidores de la PCR (como los ácidos húmicos y fúlvicos) son compuestos altamente fenólicos, lo que hace que muchos de ellos queden retenidos en la fase fenólica al inicio del proceso de extracción y sean eliminados eficientemente (O'Rourke et al., 2000). Sin embargo, el protocolo S-B, a pesar de su menor eficiencia en la eliminación de inhibidores (ver panel **c** de la **Figura D 1**) permitió recuperar resultados replicables en la misma proporción que el protocolo basado en el empleo de Fenol/Cloroformo (ver panel **b** de la **Figura D 1**). Este efecto probablemente se deba al *kit* empleado para llevar a cabo la amplificación (*Multiplex PCR Kit* de *Qiagen*), que enmascara el posible efecto inhibitorio de los extractos, puesto que su eficiencia de amplificación no se ve afectada por la presencia de estas moléculas.

En el presente trabajo se emplearon dos tipos de sales para aislar el ADN dentro del protocolo S-B, el Cloruro Sódico (NaCl) y el Tiocianato de Guanidina (GuSCN). Como puede verse en el panel **d** de la **Figura D 1**, el NaCl es menos eficiente en la eliminación de inhibidores que el GuSCN. En este último caso, el porcentaje de eliminación de inhibidores se aproxima al del método P/C. Tal y como los propios autores reconocen en su artículo, el GuSCN es más eficiente en la eliminación de moléculas inhibitorias de la PCR y su empleo está recomendado en aquellas muestras que se sospeche puedan presentar un gran contenido de inhibidores (Rohland y Hofreiter, 2007 b).

Sin embargo, la eficiencia de las dos sales en la obtención de resultados es similar, aunque menor que la del protocolo P/C. Este efecto se podría explicar debido a la elevada eficiencia del NaCl en el aislamiento de ADN (Rohland y Hofreiter, 2007 b). La razón por la que el método S-B no consigue alcanzar la eficiencia del P/C puede ser debida a que la sílica es un potente inhibidor de la PCR (O'Rourke et al., 2000). Una incompleta eliminación de la misma podría ser el motivo principal de las diferencias en la eficiencia de obtención de ADN endógeno entre ambos métodos.

A pesar de la mejor eficiencia del protocolo P/C en la obtención de resultados replicables y en la eliminación de inhibidores de la PCR, se consideró que el protocolo desarrollado por Rohland y Hofreiter (S-B) con GuSCN era el más adecuado para el estudio de muestras antiguas, puesto que presenta las siguientes ventajas técnicas:

- Evita la exposición a sustancias altamente tóxicas volátiles como el fenol y el cloroformo. Para la extracción de muestras antiguas se suele emplear una cantidad de estos reactivos proporcionales al volumen de muestra. En consecuencia, se trabaja de manera prolongada con volúmenes elevados (entre 5 y 10 ml) de estas sustancias. Además hay que tener en cuenta que la sala de extracción no se puede ventilar debido a las medidas de prevención de la contaminación. Tampoco el empleo de mascarillas con filtro para solventes orgánicos consigue proteger completamente de los vapores. El GuSCN es también una sustancia tóxica, sin embargo no es volátil por lo que las medidas de protección empleadas en los laboratorios de extracción de aDNA son suficientes.
- Aunque ambos métodos necesiten un tiempo total de dos días para completar la extracción de ADN, el protocolo S-B implica fases más cortas que permiten solapar diferentes tandas de extracción.
- El protocolo S-B reduce la manipulación de la muestra por el investigador durante las diferentes fases de la extracción, minimizando la contaminación con ADN exógeno. Por un lado, las fases de lisis con proteinasa K y descalcificación con EDTA se llevan a cabo simultáneamente, mientras que en el caso del P/C son dos pasos independientes. Por otro lado, tras la fase de digestión con proteinasa K, se emplean volúmenes más reducidos. En cambio en el protocolo P/C los volúmenes se reducen solo en el último paso de la extracción: la concentración.
- Las probabilidades de contaminación también se ven reducidas debido al empleo de material estéril en todo momento. Esto no es posible en el caso del protocolo P/C, puesto que los materiales plásticos no resisten a la acción de los solventes orgánicos.

Por consecuencia es necesario emplear material de vidrio previamente lavado, autoclavado y esterilizado manualmente.

- Muchas de las soluciones empleadas en el protocolo S-B son filtrables, permitiendo un mayor control de la contaminación de fábrica.
- En el caso de muestras bien preservadas y supuestamente con una baja concentración de moléculas inhibitoras (por ejemplo, muestras enterradas en sarcófagos que no han estado en contacto con el suelo), el GuSCN puede ser sustituido por NaCl, abaratando considerablemente el coste del protocolo.
- El protocolo S-B es escalable, permitiendo adaptar los volúmenes empleados a la cantidad de muestra inicial.
- En el protocolo S-B, la unión del ADN a la matriz de sílica (dióxido de silice) ocurre durante una incubación de tres horas. Por el contrario, en los kits comerciales de extracción, la adsorción del ADN únicamente se produce durante el momento en que el volumen pasa a través de la membrana.

Gracias a este estudio comparativo ha sido posible implementar el protocolo S-B en el laboratorio y demostrar su eficiencia en otras muestras antiguas, analizadas a lo largo de la presente tesis doctoral.

D.1.2 COMPARACIÓN DE DIFERENTES METODOLOGÍAS PARA LA AMPLIFICACIÓN DE ADN (ARTÍCULOS 2-5)

La amplificación de material genético antiguo, al igual que su extracción, representa una fase crítica, puesto que la PCR es una técnica muy sensible a la contaminación. Ello se debe a que el ADN contaminante no está degradado como el ADN antiguo, por lo que resulta amplificado preferentemente. La contaminación durante la PCR puede proceder tanto del investigador como del entorno de trabajo, por contaminación de

las superficies, reactivos o por *carry-over* de moléculas de anteriores amplificaciones que flotan en el ambiente.

Durante el desarrollo de la presente tesis doctoral, se evaluaron diferentes estrategias de amplificación. Para su discusión, en la **Tabla D 1** se reúnen todos los conjuntos muestrales analizados en los **ARTÍCULOS 3, 4 y 5** con el porcentaje correspondiente de resultados validados de acuerdo con los diferentes criterios de autenticidad. Los resultados se discuten en función del marcador analizado: ADN nuclear y ADN mitocondrial.

Tabla D 1 Porcentaje de éxito en el análisis de las muestras validadas en los **ARTÍCULOS 3, 4 y 5** en función del periodo analizado. Se aporta también el porcentaje de inhibición de los extractos en los yacimientos en los que fue testado. Debajo de cada periodo se proporciona la media del éxito de amplificación de los diferentes yacimientos incluidos y los marcadores analizados.

Periodo	Yacimiento	Época	Datación	Marcadores analizados	Nº individuos analizados	Nº individuos validados	% Éxito	% Inhibición
Medievales (nuDNA: 100,00 %)	San Esteban de Cuéllar (Segovia, España)	Edad Media	Siglo XV d.C.	nuDNA (MinFiler Kit)	7	7	100,00 %	No testado
Íberos (mtDNA: 50,00 %)	Tossal Sant Miquel (Valencia, España)	Edad del Hierro	Siglo IV-III a.C.	mtDNA (Multiplex PCR Kit)	3	1	33,33 %	33,33 %
	Los Villares (Castellón, España)	Edad del Hierro	Siglo IV-III a.C.	mtDNA (Multiplex PCR Kit)	1	1	100,00 %	No testado
	Puig de la Nau (Castellón, España)	Edad del Hierro	Siglo V a.C.	mtDNA (Multiplex PCR Kit)	12	2	16,67 %	66,67 %
Edad de los Metales (mtDNA: 63,13 %)	Maçanet (Castellón, España)	Edad del Bronce	1.500-1000 a.-C.	mtDNA (Multiplex PCR Kit)	1	1	100,00 %	No testado
	Mas d'Abad (Castellón, España)	Edad del Bronce	1.500-1000 a.-C.	mtDNA (Multiplex PCR Kit)	3	1	33,33 %	100,00 %
	Sepulcre Segond de la Costa (Castellón, España)	Calcolítico	2500-2000 a.C.	mtDNA (Multiplex PCR Kit)	5	2	40,00 %	36,36 %
	Torre d'en Domenèch (Castellón, España)	Calcolítico	2500-2000 a.C.	mtDNA (Multiplex PCR Kit)	1	1	100,00 %	No testado
	Cova de Malpaso (Castellón, España)	Calcolítico	2500-2000 a.C.	mtDNA (Multiplex PCR Kit)	4	1	25,00 %	0,00 %
	Racó de Raca (Castellón, España)	Calcolítico	2500-2000 a.C.	mtDNA (Multiplex PCR Kit)	3	1	33,33 %	No testado
	Assud de Villarreal (Castellón, España)	Calcolítico	2500-2000 a.C.	mtDNA (Multiplex PCR Kit)	1	1	100,00 %	100,00 %
Neolítico Antiguo (mtDNA: 55,66 %; nuDNA: 0 %)	Sant Pau del Camp (Barcelona, España)	Neolítico Epicardial	4.250-3.700 AP	mtDNA (Multiplex PCR Kit)	12	3	25,00 %	100,00 %
	Chaves (Huesca, España)	Neolítico Cardial	5.329-4.999 AP	mtDNA (Multiplex PCR Kit)	3	3	100,00 %	No testado
	Can Sadurní* (Barcelona, España)	Neolítico Cardial	5.475-5.305 AP	mtDNA (Multiplex PCR Kit)	Mínimo 7 (22 muestras)	7 (10 muestras)	42,00 % (en muestras)	55,56 %
				nuDNA (MiniFiler Kit)	Mínimo 5 (6 muestras)	Solo perfiles parciales y no replicables	0 %	

*Para las muestras procedentes del yacimiento de Can Sadurní no fue posible calcular los porcentajes a nivel de número de individuos porque se trataba de dientes sueltos. Por tanto se calculó a nivel de muestras.

D.1.2.1 AMPLIFICACIÓN DEL ADN NUCLEAR MEDIANTE EL KIT MINIFILER

En general los resultados obtenidos durante el desarrollo de esta tesis doctoral apuntan a que es posible obtener información genética de marcadores nucleares a partir de muestras antiguas mediante la implementación de herramientas diseñadas para el campo forense, como el *kit MiniFiler* de *Applied Biosystems* (Mulero et al., 2008). Las ventajas del empleo de este *kit* residen en el análisis de marcadores idóneos para estudiar relaciones de parentesco cercano con una batería de cebadores que amplifican fragmentos más cortos comparados con otros *kits* disponibles en el mercado, solventando así el problema de la fragmentación del material genético degradado. Otra ventaja de este *kit* es que los nueve STRs tetraméricos que amplifica están incluidos en la batería de los marcadores comúnmente empleada por los laboratorios forenses. Gracias a ello, se maneja un conjunto de marcadores para los que existe un marco de referencia a nivel poblacional actual, en el que se conocen las frecuencias alélicas y las tasas mutacionales. Además, el *kit MiniFiler* implica una reducida manipulación de la muestra: solo hay que añadir el extracto a una solución preparada y estéril. En el caso de diseños no comerciales, la reacción *multiplex* se prepara añadiendo todos los reactivos necesarios (desde la preparación de la mezcla de PCR hasta la inclusión de diferentes parejas de cebadores), aumentando la manipulación y, en consecuencia, las probabilidades de contaminación durante la preparación de la mezcla de reacción. Sin embargo, el empleo de este *kit*, y en general de todos los preparados comerciales, tiene la desventaja de que su composición es desconocida y no se puede modificar para adaptarla a las muestras en estudio.

Por el momento, solo se han conseguido analizar satisfactoriamente con este *kit* los especímenes antiguos bien preservados, como las muestras de la Iglesia de Cuéllar (100 % éxito, ver **Tabla D 1**). Estas muestras procedían de un conjunto de sepulcros de piedra localizados en el interior de la Iglesia de San Esteban de Cuéllar, habiéndose preservado en un entorno seco, que favorece la conservación del ADN (ver **Apartado I.3.3.4 y Material y Métodos del ARTÍCULO 5**). Por otro lado, el *kit MiniFiler* no permitió recuperar información consistente en el caso de muestras muy degradadas, como las muestras neolíticas de Can Sadurní (ver **Tabla D 1**). Estos resultados apuntan

a la escasez y degradación del ADN nuclear en estas muestras, debido probablemente a la deposición prolongada de los restos, que duró varios milenios. Tampoco fue posible recuperar información a partir de muestras más recientes (siglo XVIII) procedentes del osario del convento de las Madres Capuchinas de Murcia, que se preservaron en un entorno muy húmedo, desfavorable para la conservación de ADN (ver **Material y Métodos del ARTÍCULO 2**).

Para conjuntos de muestras degradadas (como las de Can Sadurní y las de Murcia) la dificultad en la obtención de resultados también podría estar relacionada con la estrategia de amplificación: en una única reacción compiten 9 parejas de cebadores para la amplificación de marcadores nucleares. Se ha demostrado que con muestras degradadas existen estrategias más exitosas para el análisis de marcadores nucleares, que se basan en amplificaciones en *multiplex* diseñadas *ad hoc*, como en el trabajo de Haak y colaboradores (Haak et al., 2010). En este caso fue posible amplificar y validar los resultados de 22 SNPs del mtDNA para todas las muestras estudiadas (21) y 25 SNPs del cromosoma Y en 3 de las 8 muestras identificadas como masculinas (37,5 %), todas previamente analizadas para el mtDNA. La principal ventaja de este diseño experimental reside en la longitud de los amplicones, lo suficientemente cortos como para poder obtener resultados en muestras degradadas que contienen ADN fragmentado.

Asimismo, es posible que la ineficiencia de este *kit* en muestras degradadas esté relacionada con la presencia de moléculas inhibidoras, que representan el 55,56 % en las muestras de Can Sadurní (ver **Tabla D 1**). Aunque la incidencia de estas moléculas no se haya estimado en las muestras de Murcia, la coloración oscura de las muestras analizadas podría apuntar a la presencia de moléculas inhibidoras, cuya infiltración habría sido facilitada por el entorno húmedo de preservación.

Las muestras neolíticas procedentes del yacimiento de Can Sadurní se analizaron también para mtDNA (ver **Material y Métodos del ARTÍCULO 3 y Tabla D 1**), dando resultados consistentes en el 42 % de las muestras estudiadas. Esto demuestra la mayor persistencia del mtDNA en restos antiguos debido a su mayor abundancia, que supera la del nuDNA en al menos tres órdenes de magnitud. También apunta a la

elevada eficiencia de la estrategia de amplificación de mtDNA, a pesar de la alta incidencia de moléculas inhibidoras (55,56 %) (ver **Apartado D.1.2.2**).

En su conjunto, estos resultados confirman la importancia del ambiente de deposición de los restos en la preservación del material genético, que se hace especialmente manifiesto a la hora de analizar nuDNA (ver **Apartado D.2.2**). Estos datos están confirmados por diferentes trabajos de otros autores. El *kit MiniFiler* ha permitido el análisis exitoso de muestras antiguas (alrededor del 3.300 AP) y bien preservadas, como las momias de la familia de Tutankamon, que proporcionaron perfiles completos o parciales para 7 de los marcadores estudiados en el 100 % de las muestras incluidas en el estudio (Hawass et al., 2010); así como en el caso de las muestras de San Esteban de Cuéllar (ver **Tabla D 1** y **Tabla 2** del **ARTÍCULO 5**). En el caso de otro conjunto de muestras medievales (del siglo VII) procedentes de los pirineos Aragoneses (España), el *kit MiniFiler* y otros *kits* forenses han permitido la recuperación de perfiles casi completos en 3 de las 7 muestras estudiadas, con un porcentaje de éxito del 43 %, (Núñez et al., 2011), calculado como el número de individuos tipados sobre el total. Este valor se sitúa por debajo del obtenido en las muestras medievales analizadas en la presente tesis doctoral (100 %). Estas similitudes y diferencias en los porcentajes de éxito parecen confirmar la importancia del estado de preservación de las muestras y más concretamente sugieren que la momificación natural o artificial ayuda a la preservación del material genético.

D.1.2.2 AMPLIFICACIÓN DEL ADN MITOCONDRIAL MEDIANTE EL *KIT MULTIPLEX* DE *QIAGEN*

El análisis de mtDNA de muestras prehistóricas proporcionó porcentajes de éxito variables en función del conjunto de muestras analizadas, los cuáles oscilan entre el 16,67 % y el 100 % (ver **Tabla D 1**). Es interesante notar como los porcentajes de éxito asociados a los diferentes períodos prehistóricos no guardan relación con su antigüedad. Además, dentro de cada uno de los períodos estudiados se detectan amplias oscilaciones en función del yacimiento. Estos resultados apuntan otra vez a

que el éxito en los análisis genéticos depende principalmente de factores no relacionados directamente con su antigüedad, como el estado de preservación de las muestras (ver **Apartado D.2.1**). Es posible corroborar esta hipótesis comparando los datos de eficiencia en la amplificación del mtDNA con los disponibles en la literatura para períodos y/o regiones parecidas a las estudiadas en la presente tesis doctoral. El Neolítico Medio de la provincia de Barcelona ha sido previamente analizado por otros autores (Sampietro et al., 2007), que revelaron un porcentaje de éxito del 47,83 % (número de individuos analizados sobre el total), similar al obtenido para las muestras neolíticas antiguas de Can Sadurní (42,00 %) pero más elevado que el de Sant Pau del Camp (25,00 %, Neolítico Antiguo Final). La eficiencia diferencial de las muestras de Sant Pau del Camp probablemente se deba a su mayor degradación, a la presencia de inhibidores (66,67 %) y a la elevada manipulación de las mismas previamente al estudio genético (ver **Apartado D.2.1**).

Por otra parte, se registra una mayor diferencia dentro de los períodos más recientes, como en el caso de las muestras íberas, objeto de estudio de esta tesis doctoral (50,00 % de éxito, ver **Tabla D 1**) y de otros autores (85,00 % de éxito) (Sampietro et al., 2005), en ambos casos procedentes del noreste de la Península Ibérica. En el caso de los datos descritos en la **Tabla D 1**, es posible ver que existe un yacimiento que proporcionó un porcentaje de resultados validados muy reducido (Puig de la Nau, 16,67 %), que disminuye el porcentaje global del período correspondiente. Este bajo porcentaje probablemente se deba a que el conjunto de muestras incluía solamente piezas óseas, por lo general bastante deterioradas a nivel macroscópico.

También hay que tener en cuenta que los porcentajes de éxito están relacionados con los resultados obtenidos durante la fase de amplificación, que primaba aquellas muestras que daban resultados consistentes desde las primeras etapas y que fueron las que se validaron posteriormente. Para la amplificación del mtDNA se utilizó en el laboratorio un *kit* de amplificación muy sensible a la presencia de bajas cantidades de ADN: el *Multiplex PCR Kit* de *Qiagen*. Durante el desarrollo de la presente tesis doctoral se empleó desde el principio el *kit* citado (**ARTÍCULOS 1, 3 y 4**) y además fue posible destacar la elevada eficiencia del mismo incluso en presencia de cantidades

considerables de inhibidores de la PCR (ver porcentajes de éxito y de inhibición en la **Tabla D 1** y ver paneles **b** y **c** de la **Figura 1** del **ARTÍCULO 1**). De hecho, en un individuo procedente del yacimiento del Assud de Villarreal, cuyos extractos inhibieron completamente una reacción estándar de PCR, fue posible amplificar el mtDNA con el *Multiplex PCR Kit* de Qiagen.

Sin embargo, la principal desventaja asociada al empleo de este *kit* también está relacionada con su sensibilidad que, así como facilita la amplificación de bajas concentraciones de ADN, también favorece la amplificación de moléculas contaminantes. Esto ocurre no solo en el caso de muestras altamente contaminadas, sino también en el caso de muestras muy degradadas o con bajas concentraciones de ADN, en las que se amplifica cualquier molécula contaminante aunque esté en muy baja concentración. Esta característica se hizo patente en el caso de las muestras de ADN extraídas con un método poco eficiente, el protocolo Q (**ARTÍCULO 1**, ver **Apartado D.1.1**). En este caso la elevada eficiencia de amplificación estaba asociada esencialmente a moléculas contaminantes introducidas durante la fase de extracción, que implicaba mucha manipulación por parte del investigador (ver **Apartado D.1.1**).

Este fenómeno también está corroborado por el análisis de los alineamientos de secuencias clonadas (ver **Tabla S7** del **ARTÍCULO 3**), donde fue posible identificar moléculas contaminantes además de la secuencia consenso en la mayoría de las muestras. En este caso, la presencia de contaminaciones se debe principalmente a la elevada degradación del ADN contenido en muestras muy antiguas, lo que resulta confirmado por la presencia de una concentración elevada de daño molecular (ver **Tabla S7** del **ARTÍCULO 3**). En este artículo, se decidió mantener y marcar las secuencias contaminadas por el personal o por *carry-over* y aquellas afectadas por el fenómeno de *Jumping-PCR* para proporcionar al lector todos los datos obtenidos y su interpretación. En la mayoría de las publicaciones de ADN antiguo, cuando estas secuencias son identificadas, se eliminan de los alineamientos, sin proporcionar al lector un marco de comparación e interpretación a nivel de frecuencia de contaminaciones y origen de las mismas.

D.1.3 SELECCIÓN DE LAS MUESTRAS Y DE LA METODOLOGÍA ADECUADA EN FUNCIÓN DE LAS FINALIDADES DEL ESTUDIO (ARTÍCULOS 3-5)

Como ya se comentó en la Introducción (**Apartado I.3.3.3**), el muestreo y análisis en paralelo de dos piezas óseas o dentales es uno de los requisitos de reproducibilidad en el campo del aDNA. Sin embargo, el material disponible no siempre reúne las características deseadas. Por ejemplo, en el caso de las muestras cardiales procedentes del yacimiento de Can Sadurní, los enterramientos humanos no estaban en conexión anatómica y solo se disponía de dientes sueltos (ver **Material y Métodos del ARTÍCULO 3**). En estos casos, puesto que se trataba de muestras muy antiguas y valiosas, una vez triturada la pieza dental, se dividió el polvo en dos porciones, realizando extracciones y amplificaciones independientes a partir de cada una de ellas. La principal desventaja que conlleva este procedimiento es que es posible confundir como endógeno un resultado procedente de contaminaciones anteriores a la separación del polvo en dos porciones.

La contaminación puede ocurrir durante la deposición del resto, por ejemplo, por ADN humano procedente de otros individuos próximos. En este caso, sería difícil su identificación como contaminante, siendo solo posible esta identificación en el caso de analizar más de una muestra y siempre que la contaminación no se haya extendido también a otras piezas. Sin embargo, en este caso, al no coincidir los dos perfiles de las dos piezas habría que eliminar el resultado por falta de replicación o recurrir a una tercera muestra para decantarse por uno de los perfiles (aunque en este caso tampoco se obtendrían garantías de que no se trate de otra pieza contaminada). Sin embargo, este tipo de contaminación no es problemático en el caso de que se haya traspasado ADN entre muestras contemporáneas y el objetivo del estudio sea de tipo poblacional, puesto que en este caso la finalidad del estudio sería el análisis de un conjunto de muestras representativas de un determinado período.

El verdadero peligro es el ADN fresco exógeno que haya entrado en contacto con la muestra y que proceda del personal que ha manipulado la misma hasta el momento en el que las dos porciones de polvo son separadas (desde su excavación, hasta su

conservación, su análisis antropológico o durante las primeras fases de limpieza y trituración de la muestra en el laboratorio). Por este motivo, cuando solo se trabaja con una muestra de un individuo, resulta imprescindible conocer todos los perfiles de todos los manipuladores para poder detectar esta contaminación.

En el caso de la presente tesis doctoral ha sido posible descartar aquellas muestras contaminadas gracias al tipaje de todos los manipuladores. En las muestras de Can Sadurní ha sido posible detectar en varias ocasiones contaminaciones por parte del personal arqueológico, antropológico o del laboratorio que manipuló las muestras en las fases iniciales (18,18 % de las muestras analizadas). En todos los casos las muestras contaminadas fueron eliminadas del análisis. Sin embargo, en la mayoría de las muestras (80 %) no se detectó ninguna contaminación puntual ni del investigador ni de otros manipuladores hasta que se analizaron los alineamientos de amplificaciones clonadas. Estas secuencias no se consideraron para el establecimiento del haplotipo consenso (ver **Apartado D.3.1.3**).

En el yacimiento neolítico de Can Sadurní, para no duplicar información genética procedente de un mismo individuo, se adoptó un criterio de reducción analizando conjuntamente la información antropológica y genética. Así, dos muestras con el mismo haplotipo mitocondrial se consideraron pertenecientes a individuos diferentes solo cuando se hubiesen determinado edades diferentes mediante criterios antropológicos o cuando se tratase de la misma pieza dental. De esta manera, las 10 muestras analizadas, se consideraron procedentes de 7 individuos, agrupando según estos criterios aquellas muestras con el mismo haplotipo (muestras CSA05-CSA11 y CSA15-CSA22-CSA23). En cambio la muestra CSA09, aunque presentaba el mismo haplotipo que las CSA05 y CSA11 (procedentes de individuos adultos), fue considerada separadamente porque se trataba de un espécimen infantil.

Si la agrupación no se correspondiera con la realidad, estaríamos subestimando las frecuencias de los haplotipos y haplogrupos correspondientes y a la vez sobrestimando la diversidad genética de la población.



Por otra parte, para llevar a cabo el análisis genético de un conjunto de muestras antiguas es necesario seguir determinadas pautas que afectan tanto al muestreo como a la elección del marcador genético a estudiar, y que varían en función de los objetivos planteados. En el caso de la presente tesis doctoral, se han incluido dos aplicaciones diferentes: (1) el análisis de relaciones familiares y (2) el estudio diacrónico de una población.

En el primer caso –análisis de relaciones familiares–, se estudió un conjunto de sepulcros medievales, que según el registro arqueológico e histórico podrían contener individuos relacionados (ver **Hipótesis Arqueológicas en Material y Métodos del ARTÍCULO 5**). El análisis genético en este caso tendría la finalidad de comprobar dichas hipótesis. En lo que respecta a la selección de la muestra para este tipo de estudios, únicamente es necesario que todos los individuos estén representados. Además, para garantizar la reproducibilidad de los resultados, la selección de muestras tenía que incluir al menos dos piezas óseas o dentales de todos los individuos que se querían estudiar.

En cuanto al marcador empleado, en este caso, al ser el objetivo del estudio la identificación y el establecimiento de relaciones familiares, el análisis genético se centró en marcadores altamente polimórficos, de herencia diploide, como los del ADN nuclear. El éxito en la amplificación de estos marcadores permitió determinar relaciones de paternidad, maternidad o hermandad, en algunos casos con probabilidades solo ligeramente inferiores a las empleadas hoy en día para muestras forenses (ver **Apartado D.5**). Sin embargo, si se analizaran marcadores de linaje, como el mtDNA o el cromosoma Y, solo sería posible excluir relaciones a lo largo de la línea materna o paterna en el caso de haplotipos diferentes. La presencia de haplotipos coincidentes simplemente podría indicar una relación a lo largo de un linaje, que podría ser cercana o lejana, pero el resultado no sería tan concluyente como en el caso de los marcadores nucleares.

En el segundo caso –estudio diacrónico-poblacional–, se analizó un conjunto de muestras de diferentes cronologías, todas ellas procedentes del noreste de la Península Ibérica (ver **Material y Métodos** de los **ARTÍCULOS 3 y 4**). Estos trabajos tenían dos finalidades: estudiar el impacto de los primeros inmigrantes neolíticos que llegaron a la Península Ibérica (cultura Cardial, 7.900-6.400 AP) y evaluar la continuidad o discontinuidad genética entre los sucesivos pobladores de la misma región. En este caso, además de incluir al menos dos piezas óseas o dentales de todos los individuos que se querían estudiar, la selección de muestras debía cumplir con el requisito de representatividad. Es decir, cada uno de los períodos estudiados debía incluir un número suficiente de muestras de diferentes individuos no relacionados por vía materna para que éstas fueran representativas de la época correspondiente.

Teniendo en cuenta las dificultades asociadas al estudio poblacional de muestras antiguas, como la escasez de la muestra, su difícil acceso, y la ausencia en muchos casos de información relativa al parentesco, resulta muy complicado cumplir con estos requisitos. Las muestras arqueológicas representan un material valioso y a veces muy escaso, sobre todo en lo que respecta a épocas muy antiguas. A efectos comparativos, en los estudios de genética de poblaciones actuales se selecciona un amplio número de muestras aleatorias, que proceden de individuos no relacionados de la población autóctona (con los cuatros abuelos pertenecientes a esa misma población).

Como ya se ha comentado en el **Apartado D.1.2**, existen porcentajes de éxito diferenciales dentro de un mismo yacimiento, que se deben al diferente estado de preservación, nivel de absorción de moléculas inhibidoras y posible contaminación de cada uno de los individuos y muestras estudiadas dentro de un mismo yacimiento. En ese mismo apartado también se ha podido destacar la mejor o peor preservación de los restos a nivel de yacimientos contemporáneos. Ambos factores implican la recolección de un conjunto de datos sesgados y una posible sobrerrepresentación en el caso de los individuos y yacimientos asociados a mejor preservación. Sin embargo, en el caso del presente trabajo es posible destacar que los yacimientos asociados a elevadas eficiencias (100 %) no incluían un elevado número de individuos (en dos casos un único individuo y en otro caso tres, ver **Tabla D 1**).

A la hora de analizar un conjunto de muestras representativas de una época también hay que evitar el estudio de individuos enterrados conjuntamente, puesto que podrían seguir un patrón de enterramiento dictado por relaciones familiares.

Las únicas muestras que presentan haplotipos mitocondriales coincidentes y que proceden del mismo yacimiento, a parte del caso citado de Can Sadurní, son las que proceden de Puig de la Nau (ver **Tabla D 2**). En este caso se trataba de individuos enterrados separadamente, por esta razón se incluyeron en el análisis poblacional. Sin embargo, esto no excluye que no pudieran tener una relación a lo largo de la línea materna y por lo tanto que estemos sobrestimando la frecuencia del haplotipo y del haplogrupo correspondientes.

Asimismo, para aumentar la representatividad de un conjunto de individuos es mejor analizar muestras procedentes de diferentes yacimientos de la misma época. Por ello, para cada período prehistórico analizado en la presente tesis doctoral se han estudiado conjuntos muestrales procedentes de diferentes yacimientos (mínimo 3 por cada periodo, ver **Apartado D.4**). Sin embargo, es importante hacer notar que la interpretación de movimientos poblacionales a gran escala, como el mecanismo de difusión del Neolítico, se ha basado en otros estudios en un conjunto de individuos procedentes de un solo yacimiento. Este es el caso del trabajo de Sampietro y colaboradores (2007) en el que se analizaron 11 individuos de un mismo yacimiento (Camí de Can Grau) del Neolítico Medio. En otro trabajo más reciente (Lacan et al., 2011), se presentan los resultados de 29 individuos del Neolítico Final procedentes del yacimiento francés de Treilles. En este caso se analizaron diferentes marcadores (mtDNA, STRs autosómicos y cromosoma Y) y se evaluaron las posibles relaciones familiares entre los individuos de la necrópolis. Aunque en el texto principal no se haga referencia a ello, en el material suplementario se hace hincapié en la baja representatividad de la muestra estudiada debido a que procede de una sola necrópolis. El análisis a nivel poblacional se justifica como búsqueda de afinidades entre la población de Treilles y las poblaciones actuales.

Para todas las muestras prehistóricas analizadas, teniendo en cuenta las finalidades de carácter poblacional del estudio, se analizó el mtDNA, ya que se trata de un marcador que permite seguir un linaje a lo largo de diversas generaciones, y su preservación es más probable en un entorno arqueológico (ver **Apartado I.2**). Sin embargo, también se realizaron algunos ensayos para comprobar si era posible obtener ADN nuclear a partir de algunas muestras neolíticas (Can Sadurní, ver **Tabla D 1**). La obtención de perfiles parciales y el fracaso en la replicación de los experimentos confirmaron la elevada degradación de estas muestras, no aptas para análisis de ADN nuclear. Los pocos sistemas tipados presentaban además una reducida altura de los picos en los electroferogramas y no existía una relación entre la longitud del sistema amplificado y el éxito en la amplificación. Todo ello, junto con el hecho de que se detectaron ciertos contaminantes traza (en muy baja concentración y difícilmente detectables) en los controles negativos, hace pensar que se tratara de artefactos o contaminaciones.

D.2 Evaluación de la posibilidad de recuperar ADN a partir de muestras antiguas en función de diferentes variables como son el estado de preservación, la procedencia, la antigüedad, las características morfológicas del tejido y el contenido de ADN en los extractos estimado mediante métodos espectrofotométricos.

En las primeras fases del desarrollo de esta tesis doctoral, se evaluaron diferentes características que podrían influir en el éxito en la recuperación de ADN endógeno a partir de muestras antiguas. Éstas se detallan a continuación.

D.2.1 PROCEDENCIA Y ANTIGÜEDAD DE LAS MUESTRAS (ARTÍCULOS 1, 3 Y 4)

Mediante el estudio prospectivo llevado a cabo al principio de la presente tesis doctoral fue posible detectar cierta influencia del entorno de enterramiento y de la localización geográfica (a nivel de yacimiento y región) sobre la eficiencia de amplificación (ver **Figura D 1**). En lo que respecta a la antigüedad, no fue posible detectar diferencias significativas (ver **Tabla A1 del Anexo del ARTÍCULO 1**).

Estas observaciones están en concordancia con los resultados de otros autores, que afirman que las condiciones a las que está sometida la muestra después de su deposición influyen considerablemente sobre la preservación del material genético (Lorenz y Wackernagel, 1987; Romanowski et al., 1991; Alvarez et al., 1998; Burger et al., 1999; de Torres et al., 2002).

En trabajos posteriores ha sido posible confirmar esta hipótesis, puesto que se detectó una elevada variabilidad entre la eficiencia en la obtención de resultados de diferentes yacimientos con dataciones similares procedentes de la misma región (ver **Tabla D 1 en el Apartado D.1.1**). Así en el caso de las muestras neolíticas antiguas, se obtuvieron porcentajes de éxito muy diferentes entre los tres yacimientos estudiados, todos ellos procedentes del noreste de la Península Ibérica (Aragón y Cataluña). Para Chaves y San Pau del Camp se tiparon de forma reproducible el 100 % y el 25 % de los individuos estudiados, respectivamente. Para Can Sadurní, el 42 % del total de las muestras. Una

razón que podría explicar el bajo porcentaje de éxito en la obtención de resultados a partir de las muestras procedentes de Sant Pau del Camp es que se trata de un yacimiento al aire libre, mientras que Can Sadurní y Chaves están emplazados en cuevas. Este segundo ambiente de deposición proporciona temperaturas más estables a lo largo del tiempo y una mayor protección frente a factores medioambientales, como la precipitación (Hedges y Millard, 1995; Bollongino et al., 2008). Así las muestras procedentes de Sant Pau del Camp probablemente contenían material genético más degradado, puesto que estas condiciones favorables no se cumplen en el caso de un yacimiento al aire libre (ver **Apartado I.3.3.4**). Además, las muestras de Sant Pau del Camp presentaban importantes depósitos calcáreos, que fueron eliminados por el personal antropológico empleando ácido acético (Estebaranz et al., 2008). Su posible penetración en las muestras, y la consecuente degradación del material genético y/o la inhibición de la PCR, probablemente también hayan determinado el reducido porcentaje de éxito. El tratamiento con ácido acético se identificó como la causa principal de la ausencia de resultados de un estudio preliminar realizado con otras muestras del mismo yacimiento (Estebaranz et al., 2008).

Se destacan diferencias también en muestras más recientes de un mismo período cronológico. Por ejemplo, en el caso de las muestras íberas se obtuvieron porcentajes de éxito muy reducidos en el yacimiento de Puig de la Nau (16,67 %, ver **Tabla D 1** en el **Apartado D.1.1**). Como ya se ha comentado en el **Apartado D.1.2.2**, la baja eficiencia de estas muestras podría estar relacionada con su elevada degradación o con la presencia de inhibidores, detectados en el 68 % de las muestras (ver **Tabla D 1**). Desde el punto de vista de su localización, merece la pena destacar que se trata de un yacimiento situado a las orillas del mar mediterráneo, mientras que los otros dos yacimientos de la misma época, Tossal de Sant Miquel y Los Villares, se encuentran respectivamente a unos 30 y 80 Km de la costa. La localización de Puig de la Nau más próxima al mar podría haber influido en la preservación de las muestras debido a un ambiente más húmedo, que por un lado perjudica la preservación del ADN y, por otro lado, podría haber aumentado la permeabilidad a las moléculas inhibitoras en el

interior de los huesos, más porosos que los dientes (Kurosaki et al., 1993; Oota et al., 1995).

En el caso de las muestras de la Edad de los Metales, resulta sorprendente que el yacimiento que proporcionó una eficiencia menor de resultados (25 %) estuviese emplazado en una cueva (Cova de Malpaso, ver **Tabla D 1** en el **Apartado D.1.1**). Como se ha comentado anteriormente, las cuevas preservan las muestras de los cambios de temperatura drásticos y mantienen unas condiciones climáticas estables. La razón por la que no se obtuvo un porcentaje de éxito mayor para los individuos procedentes de la Cova de Malpaso, puede deberse a que se trataba de restos procedentes de una excavación muy antigua. Los restos se exhumaron en los años 40 (Díaz, 2002) y se conservaron a temperatura ambiente durante más de 60 años en el Museo de Prehistoria de la Diputación Provincial de Valencia, antes de su análisis genético. Como apuntan otros autores (Pruvost et al., 2007), la preservación de los restos después de la excavación juega un papel importante, ya que las muestras recién excavadas tienen más probabilidades de proporcionar resultados positivos, puesto que durante el almacenaje están expuestas a temperaturas más elevadas y a una mayor oxigenación, factores ambos que aceleran la degradación del ADN (Pruvost et al., 2007).

Sin embargo, para la interpretación de estos resultados hay que tener en cuenta el tamaño muestral analizado. Para los conjuntos de muestras calcolíticas el número de individuos analizados por yacimiento era siempre inferior a 5 (ver **Tabla D 1**). Por esta razón, los porcentajes de éxito calculados sobre un conjunto tan reducido podría sesgar la interpretación de los resultados. Para confirmar las hipótesis aquí planteadas sería necesario ampliar el conjunto muestral de cada yacimiento.

D.2.2 CARACTERÍSTICAS DE LAS MUESTRAS (ARTÍCULOS 1 Y 5)

En el **ARTÍCULO 1** se evaluó la relación existente entre las características macroscópicas de la muestra (integridad y color) y el tipo de tejido analizado (diente o hueso) con la eficiencia de obtención de resultados y la presencia de inhibidores.

Respecto al tipo de pieza (dental u ósea), se encontraron diferencias significativas a nivel de amplificación y secuenciación del mtDNA, pero no respecto a la inhibición ni a la replicación de los resultados (ver **Tabla A2 del Anexo del ARTÍCULO 1**). Cabía esperar que los resultados apuntaran a una influencia significativa del tipo de tejido empleado (hueso o diente) también sobre la obtención de resultados replicables y sobre todo a nivel de inhibición. En la literatura se afirma que el diente es la pieza que ofrece mayores garantías a la hora de recuperar información genética endógena, sobre todo por la protección proporcionada por el esmalte y por su reducida porosidad respecto a los huesos (Kurosaki et al., 1993; Oota et al., 1995). Concretamente, se ha demostrado que el ADN puede permanecer en el interior de los dientes en condiciones extremas de humedad, temperatura y acidez (Schwartz et al., 1991; Alvarez García et al., 1996).

El hecho de que no haya sido posible detectar diferencias significativas entre los dos tipos de muestra respecto a la eficiencia de replicación probablemente está relacionado con dos factores: la disparidad de tamaño muestral entre los dos tipos de tejido (69 muestras dentales y 21 óseas) y la selección no aleatoria de la mayoría de muestras óseas (61 %, 9 sobre 21), ya que únicamente se analizaron aquellas que dieron resultados positivos en análisis precedentes al desarrollo de la presente tesis doctoral. Esto se refleja a nivel del porcentaje de obtención de secuencias no contaminadas, que muestra diferencias significativas debido a la mejor eficiencia de los huesos respecto a los dientes (38 y 28 % de las secuencias obtenidas respectivamente).

En lo que concierne a la inhibición, los dos tipos de pieza no presentaban diferencias: alrededor del 40 % de extractos de huesos y dientes mostraron poder inhibitor. La explicación de la falta de diferencias significativas a este nivel, probablemente también esté relacionada con el sesgo introducido a la hora de muestrear estas piezas (ver párrafo anterior).

Sin embargo, un estudio realizado con posterioridad a lo largo de la presente tesis doctoral sí que permitió detectar diferencias sustanciales entre hueso y diente a nivel de marcadores nucleares. Se trata del estudio de relaciones familiares de la necrópolis

de San Esteban de Cuéllar (ver **Material y Métodos del ARTÍCULO 5**), en el que para un individuo adulto (7SC), se analizaron dos dientes (7SC1, 7SC2) y un fragmento de hueso correspondiente a una costilla (7SC3) (ver **Tablas 1 y 2 del ARTÍCULO 5**). Las dos muestras dentales proporcionaron perfiles completos para los 9 marcadores en todas las amplificaciones realizadas. En cambio, las dos amplificaciones de la muestra ósea permitieron tipar en el primer caso 7 marcadores, de los cuales dos presentaron *allelic dropout* y en el segundo caso un solo marcador con *allelic dropout* (ver **Apartado D.5.1**). Aunque se trata de un solo caso, estos resultados sugieren una mejor preservación del ADN en tejidos dentarios que en óseos.

Las características macroscópicas de integridad y coloración de la muestra se suelen relacionar con el estado de degradación de la misma y con la presencia de moléculas inhibitoras (Kolman y Tuross, 2000). Una coloración oscura se suele asociar con una elevada degradación de la muestra y la posible infiltración de sustancias del entorno de deposición, que pueden incluir moléculas inhibitoras (O'Rourke et al., 2000). Cabría esperar que aquellas piezas íntegras y de coloración clara contuvieran menos inhibidores y permitieran la obtención de resultados replicables. Por un lado, las piezas rotas están menos protegidas contra la degradación y más sujetas a la infiltración de moléculas contaminantes. Por otro lado, en la literatura la falta de integridad suele asociarse a un elevado poder inhibitor, debido a una mayor facilidad de infiltración de estas moléculas, y por consecuencia a una menor eficiencia de amplificación (Pääbo et al., 1989; Hagelberg y Clegg, 1991; Kolman y Tuross, 2000). Sin embargo, el conjunto de datos analizados incluía principalmente muestras íntegras y sin fisuras (74 %, correspondiente a 67 muestras). Probablemente esta es la razón por la que no ha sido posible detectar diferencias significativas con muestras no íntegras.

El hecho de que la coloración oscura de las muestras no se pudiera asociar con el poder inhibitor de los extractos, podría deberse a que otros factores más relacionados con la inhibición, como el yacimiento de procedencia o protocolo empleado, pudieran estar enmascarando una asociación positiva.

D.2.3 EVALUACIÓN DE LAS MEDIDAS DE ABSORBANCIA (ARTÍCULO 1)

Aunque la absorbancia a 260 nm haya sido utilizada como medida estándar en la cuantificación de ADN (Sambrook y Russell, 2001), en otros trabajos se ha demostrado que no se trata de un método adecuado para muestras antiguas por diferentes razones. En primer lugar, la medida de absorbancia se realiza sobre el ADN total, pudiendo ser éste de la muestra o de otro origen (contaminantes bacterianos, fúngicos o ADN de los manipuladores). En segundo lugar, es posible que los extractos de ADN antiguo contengan sustancias inhibitoras que son co-purificadas junto con el material genético y que pueden sesgar las medidas de absorbancia (Lassen et al., 1994; Montiel Duarte, 2001). Por ejemplo los ácidos húmicos absorben entre 208 y 350 nm, mientras que los fúlvicos entre 208 y 400 nm (Montiel Duarte, 2001).

Tratándose de una técnica de *screening* rápida y de bajo coste, se testó inicialmente con alguna de las muestras incluidas en la presente tesis doctoral. El análisis de los resultados destacó la ausencia de relaciones significativas entre las medidas de absorbancia espectrofotométrica a 260 nm y la eficiencia en la obtención de los resultados. Una posible explicación es que la inhibición juegue un papel más importante. Así en el caso de muestras que contengan una cantidad suficiente de ADN, es posible que no sean amplificadas debido a la presencia de inhibidores en los extractos.

El espectro de absorbancia testado incluía longitudes de onda desde los 220 hasta los 350 nm. Así se pudieron calcular los ratios de absorbancia a 260/230 y 260/280, que reflejan la pureza de los extractos. Ambas medidas pueden informar respectivamente acerca de la correcta eliminación de sales (en este caso EDTA) y proteínas durante el aislamiento del ADN. El primer cociente se refiere a la proporción entre ADN y sales, que absorben a 230 nm, mientras que el segundo a la proporción entre ADN y proteínas, que absorben a 280 nm. Valores superiores a 2 para 260/230 (Yeates et al., 1998) y de alrededor de 2 para 260/280 son indicadores de pureza de los extractos (Glasel, 1995).

Resulta interesante destacar los reducidos valores del promedio (entre 0,12 y 0,7) de la relación de absorbancia a 260/230 en todos los protocolos testados (Q, S-B y P/C) que sugieren una escasa eliminación de sales o una elevada presencia de inhibidores. Este ratio es mucho más bajo en el caso del protocolo Q (0,12) que en S-B (0,56) y P/C (0,7), destacando la baja eficiencia de este protocolo en la obtención de ADN libre de sales. También en lo que concierne a la relación de absorbancia a 260/280 existen diferencias significativas entre los tres protocolos; mientras que los protocolos S-B y P/C presentan valores ligeramente bajos (respectivamente un promedio de 1,47 y 1,48), el método Q exhibe un valor muy elevado (7,39 en promedio). Los primeros dos valores indican una eficiencia parecida entre los protocolos P/C y S-B en la eliminación de proteínas, reflejado en sus eficiencias parecidas en la obtención de resultados y en la eliminación de inhibidores (ver **Apartado D.1.1**).

La razón por la que el protocolo Q proporciona valores de ratios alejados del esperado se debe a que también se alejan de los valores esperados las absorbancias a nivel de las tres longitudes de onda empleadas para el cálculo de los ratios (en promedio, 0,37 a 260 nm; 0,07 a 280 nm; 3,00 a 230 nm). Además, para la interpretación de estos resultados hay que tener en cuenta que las medidas espectrofotométricas se obtuvieron a partir de un volumen muy reducido (1,5 µl), como requerido por el protocolo específico. Esto puede haber introducido errores de medición debido a la baja cantidad de ADN en los extractos y a su heterogeneidad molecular.

Estas observaciones ponen en evidencia la escasa especificidad de este método para estimar la cantidad y calidad de los extractos de ADN antiguo. Para ello se suelen emplear métodos más específicos, como la PCR en tiempo real (Alonso et al., 2004). Sin embargo, también mediante esta técnica, aunque se esté midiendo un marcador específico, es imposible distinguir entre moléculas endógenas y contaminantes de la misma especie que, en el caso de trabajar con muestras humanas, incluirían a los manipuladores. Se presenta el mismo problema para las incipientes técnicas de secuenciación de nueva generación (NGS), mediante las cuales es posible analizar todas las moléculas presentes en un extracto y también distinguir y cuantificar la contribución de moléculas contaminantes de otras especies (Green et al., 2010; Prüfer et al., 2010). Sin embargo, para muestras humanas antiguas, la distinción entre

moléculas endógenas y contaminantes de los manipuladores solo resulta posible mediante la comparación con los perfiles de todas aquellas personas que hayan podido entrar en contacto con la muestra.

D.3 Validación de los resultados de ADN mitocondrial y nuclear obtenidos a partir de muestras históricas y prehistóricas mediante los pertinentes criterios de autenticidad, aceptados internacionalmente.

Como ya se ha insistido a lo largo de la introducción y de la discusión de la presente tesis doctoral, en el estudio de muestras antiguas se necesita una correcta validación de los resultados para confirmar su autenticidad. La razón principal reside en demostrar la ausencia de contaminación en los resultados por parte de todas las posibles fuentes.

Hay que tener en cuenta que el cumplimiento de todos y cada uno de los criterios de autenticidad representa el factor más limitante del análisis de muestras antiguas, puesto que implican un gran esfuerzo económico y un largo proceso experimental. Sin embargo, no todos los estudios de aDNA tienen las mismas problemáticas. Por ello, en la presente tesis doctoral y en las publicaciones incluidas, se intenta aportar toda la información posible acerca de los criterios de autenticidad empleados y de su interpretación, dependiendo del contexto de los resultados obtenidos y de los objetivos planteados.

A continuación se discuten los principales criterios de autenticidad adoptados.

D.3.1 DETECCIÓN Y CONTROL DE LA CONTAMINACIÓN DURANTE LA EXTRACCIÓN Y LA AMPLIFICACIÓN (ARTÍCULOS 3-5)

Gracias al correcto diseño y al estricto control de acceso de las infraestructuras reservadas al estudio de muestras antiguas, fue posible monitorizar constantemente la contaminación. Además, los controles negativos empleados en las diferentes fases del análisis permitieron identificar y eliminar todos aquellos resultados susceptibles de haber sido contaminados (ARTÍCULOS 3, 4 y 5). Solo se consideraron válidas aquellas extracciones y amplificaciones sin rastro de contaminación.

D.3.1.1 CONTAMINACIÓN EN LOS BLANCOS DE EXTRACCIÓN

Para cada extracción se amplificaron los correspondientes controles negativos por duplicado y separadamente de las muestras con las que se habían extraído, para no confundir la contaminación durante el proceso de extracción con la producida durante el proceso de PCR. En todos los casos estos resultados fueron negativos, lo que sugiere la correcta extracción de las muestras y la ausencia de moléculas contaminantes.

Sin embargo, para investigar más sobre la posible presencia de contaminantes no detectados, se llevó a cabo la amplificación repetida: en este caso una media de nueve veces por blanco, de todos los controles implicados en las extracciones de las muestras neolíticas. Los resultados de estos experimentos fueron en el 93 % de los casos negativos. Para el porcentaje restante se obtuvieron amplificaciones positivas. Estas últimas se secuenciaron con el fin de identificar el posible origen de la contaminación y se pudo comprobar que no procedía de los manipuladores. Además, no se detectó reproducibilidad de una misma contaminación para un mismo blanco de extracción, amplificado unas nueve veces en una sola tanda de PCR.

Estos resultados apuntan a una procedencia ambiental (moléculas de ADN flotante o procedente del ambiente de trabajo) de estos contaminantes, a pesar de todas las precauciones tomadas para su eliminación, como la limpieza constante de todas las superficies, la irradiación del laboratorio con luz ultravioleta, el empleo de material y reactivos estériles, etc. Probablemente este 7 % de contaminaciones se debe a la elevada sensibilidad del *kit* de amplificación empleado. Además, la obtención de diferentes haplotipos para un solo blanco de extracción evidencia la falta de reproducibilidad de estos resultados.

Este pequeño porcentaje de contaminación aleatoria, así como afectó a los blancos, es presumible que hubiera afectado a algunas de las muestras estudiadas. Sin embargo, los resultados procedentes de muestras antiguas fueron extensivamente replicados a todos los niveles (extracción, amplificación y análisis en otro laboratorio), incluida la replicación de los resultados entre diferentes amplificaciones. Por ello, se descartó la

posibilidad de confundir este pequeño porcentaje de contaminación aleatorio (7 %) con resultados endógenos a las muestras (ver **Material y Métodos del ARTÍCULO 3**).

D.3.1.2 CONTAMINACIÓN EN LOS BLANCOS DE PCR

En la **Tabla D 2** se resume la incidencia de la contaminación en los blancos de PCR en función del período y yacimiento estudiados. Como puede verse, se contaminaron un 2,54 % de las amplificaciones llevadas a cabo. Se trata de un valor reducido, sobre todo si se tiene en cuenta la elevada sensibilidad del kit de amplificación empleado para la amplificación de mtDNA (ver **Apartado D.1.2.2**). Es interesante destacar cómo los porcentajes de contaminación no siguen un orden cronológico. Aunque su mayor incidencia (7,37 %) se produce en el conjunto muestras más antiguas (neolíticas) y la más baja en las muestras medievales (0,00 %), las muestras íberas (2,38 %) presentan un porcentaje más elevado que las calcolíticas (0,90 %). También es importante destacar que los yacimientos en los que se detectó una mayor contaminación a nivel de PCR proporcionaron también porcentajes de éxito reducidos (ver **Tabla D 2 y D 1**). Este es el caso del yacimiento íbero de Puig de la Nau (contaminación de PCR: 7,14 %, porcentaje de éxito: 16,67 %) o del neolítico de Sant Pau del Camp (contaminación de PCR: 7,14 %, porcentaje de éxito: 25,00 %). Esto se debe a que durante el desarrollo de los análisis todas aquellas muestras por las que se detectaron contaminantes, en este caso en las secuencias directas, fueron descartadas.

El porcentaje de contaminación de los controles negativos de una PCR guarda relación con la manipulación, con la presencia de ADN en los extractos amplificados en paralelo, y con la presencia de contaminantes en los reactivos y en el ambiente. En lo que concierne a la manipulación, todas las muestras fueron amplificadas con *kits* (mtDNA y nuDNA) que implicaban una manipulación muy reducida y similar entre sí. También se excluye la contaminación de los reactivos, puesto que no se obtuvieron amplificaciones contaminadas relacionadas con unas alícuotas concretas de reactivos. La causa más plausible para explicar la contaminación de los blancos de PCR es, por tanto, la presencia de ADN en el ambiente de trabajo. Este ADN podría proceder de las

muestras analizadas en paralelo, de las que se evapora parte del extracto, que puede ser arrastrado con los guantes al abrir el tubo o con las pipetas. También podrían proceder del investigador o del ambiente. Dado que todas las amplificaciones de muestras antiguas con controles negativos contaminados se eliminaron, se siguió trabajando con un conjunto de muestras supuestamente no contaminadas a ningún nivel. Sin embargo, la resolución de los geles de agarosa empleados para la visualización de los resultados de las amplificaciones es reducida. Por esta razón, el hecho de que no sea posible detectar visualmente la presencia de contaminantes en los blancos no implica necesariamente su completa ausencia.

Tabla D 2 Porcentajes de amplificaciones contaminadas en relación a los períodos estudiados, el yacimiento, y el tipo de marcadores analizados. Se señala también el porcentaje de éxito en la obtención y validación de los resultados.

Período	% Contaminación PCR	Yacimiento	Marcadores analizados	Nº individuos analizados	% Contaminación PCR	% Éxito
Medievales	0,00 %	San Esteban de Cuellar (Segovia, España)	nuDNA	7	0,00 %	100,00 %
Íberos	2,38 %	Tossal Sant Miquel (Valencia, España)	mtDNA	3	0,00 %	33,33 %
		Los Villares (Castellón, España)	mtDNA	1	0,00 %	100,00 %
		Puig de la Nau (Castellón, España)	mtDNA	12	7,14 %	16,67 %
Edad de los Metales	0,90 %	Maçanet (Castellón, España)	mtDNA	1	0,00 %	100,00 %
		Mas d'Abad (Castellón, España)	mtDNA	3	0,00 %	33,33 %
		Sepulcre Segond de la Costa (Castellón, España)	mtDNA	5	0,00 %	40,00 %
		Torre d'en Domenèch (Castellón, España)	mtDNA	1	0,00 %	100,00 %
		Cova de Malpaso (Castellón, España)	mtDNA	4	1,96 %	25,00 %
		Racó de Raca (Castellón, España)	mtDNA	3	4,35 %	33,33 %
		Assud de Villarreal (Castellón, España)	mtDNA	1	0,00 %	100,00 %
Neolítico Antiguo	7,37 %	Sant Pau del Camp (Barcelona, España)	mtDNA	12	7,96 %	25,00 %
		Chaves (Huesca, España)	mtDNA	3	5,26 %	100,00 %
		Can Sadurní* (Barcelona, España)	mtDNA	Mínimo 7 (22 muestras)	8,91 %	42,00 % (en muestras)
			nuDNA	Mínimo 5 (6 muestras)	Contaminación parcial (alrededor del 40 % de los marcadores, con picos bajos)	0,00 % (en muestras)

*Para las muestras procedentes del yacimiento de Can Sadurní no fue posible calcular los porcentajes a nivel de número de individuos porque se trataba de dientes sueltos. Por tanto se calculó a nivel de muestras.

D.3.1.3 CONTAMINACIÓN NO DETECTADA EN LOS BLANCOS

A pesar del control de la contaminación mediante blancos de extracción y de amplificación, fue posible detectar moléculas contaminantes de distinta procedencia en las sucesivas fases de análisis, que van desde la obtención de una secuencia o perfil genético hasta la clonación bacteriana en el caso de las amplificaciones de ADN mitocondrial. Esto no ocurrió en el caso de las muestras medievales de San Esteban de Cuellar, en las que se analizó ADN nuclear, que no presentaron ningún rastro de

contaminación. Sin embargo, en el caso de las muestras de Can Sadurní que también se amplificaron para los mismos marcadores se detectó cierta presencia de contaminación en los controles de PCR, en forma de perfiles con picos reducidos. Probablemente se trataba de artefactos o de contaminaciones ambientales, puesto que no había coincidencia con ninguno de los posibles manipuladores.

En cuanto al mtDNA, se consideraron como válidos los resultados que presentaban blancos de extracción y amplificación limpios y que permitían la obtención de secuencias replicables mediante diferentes amplificaciones a partir de uno o más extractos. Por otro lado, se eliminaron aquellas muestras que a lo largo del análisis no proporcionaron resultados o por las que se detectaron resultados contaminados. Además, en el presente trabajo se han descartado todas aquellas muestras que presentaran el mismo haplotipo que uno de los manipuladores, puesto que resulta imposible determinar su origen endógeno o contaminante. Se consideraron como válidos solo aquellos resultados sin contaminación y replicables a nivel de diferentes amplificaciones y extracciones.

Como es posible ver en la **Tabla D 3**, se distingue entre contaminación a nivel de secuenciación directa y en la sucesiva fase de alineamiento de clones. La detección de contaminantes a nivel de secuenciación directa indica la presencia masiva de ADN exógeno. Por ello, estas muestras no se analizaron ulteriormente y solo se clonaron aquellas que no proporcionaron resultados contaminados en dicha fase. En otros laboratorios, como el de la Universidad de Burdeos (LAPP-PACEA, *Université de Bordeaux 1 – CNRS, Talence, Francia*), se procede a la clonación de todos los amplificados, sin obtener previamente su secuencia directa. Sin embargo, la secuenciación directa de los amplificados y la replicación de estos resultados como criterio indispensable para proceder a su clonación presenta la ventaja de que permite acortar los tiempos de análisis y reducir el presupuesto en el caso de muestras que presentan contaminaciones masivas o que no proporcionan resultados replicables.

Tabla D 3 Porcentajes de muestras que no proporcionaron resultados reproducibles por yacimiento y período cronológico analizado. Se distingue entre las diferentes causas de fracaso: la ausencia de amplificación y la detección de contaminantes en el análisis de secuencias directas y clones, señalándose la fuente de contaminación más probable.

Período	Yacimiento	Marcadores analizados	Nº individuos analizados	Nº individuos validados	SECUENCIACIÓN DIRECTA		CLONACIÓN	
					% muestras sin resultados (%S)	% muestras contaminadas (%C)	Detección contaminaciones en clones	% Éxito
Medievales %S: 0 %C: 0	San Esteban de Cuéllar (Segovia, España)	nuDNA	7	7	0,00 %	0,00 %	-	100,00 %
Íberos %S:19,44 %C:30,55	Tossal Sant Miquel (Valencia, España)	mtDNA	3	1	33,33 %	33,33 % (carry-over)	No	33,33 %
	Los Villares (Castellón, España)	mtDNA	1	1	0,00 %	0,00 %	sí (ambiental)	100,00 %
	Puig de la Nau (Castellón, España)	mtDNA	12	2	25,00 %	41,66 % (investigador) 16,66 % (ambiental)	Sí (investigador, ambiental)	16,67 %
Edad de los Metales %S:9,28 %C:33,80	Maçanet (Castellón, España)	mtDNA	1	1	0,00 %	0,00 %	sí (investigador)	100,00 %
	Mas d'Abad (Castellón, España)	mtDNA	3	1	0,00 %	66,66 % (ambiental)	sí (ambiental)	33,33 %
	Sepulcre Segond de la Costa (Castellón, España)	mtDNA	5	2	40,00 %	20,00 % (ambiental)	sí (investigador, ambiental)	40,00 %
	Torre d'en Domenèch (Castellón, España)	mtDNA	1	1	0,00 %	0,00 %	sí (investigador)	100,00 %
	Cova de Malpaso (Castellón, España)	mtDNA	4	1	25,00 %	50 % (investigador, ambiental)	sí (investigador)	25,00 %
	Racó de Raca (Castellón, España)	mtDNA	3	1	0,00 %	66,66 % (investigador, ambiental)	sí (investigador)	33,33 %
	Assud de Villarreal (Castellón, España)	mtDNA	1	1	0,00 %	0,00 %	sí (investigador, ambiental)	100,00 %
Neolítico Antiguo mtDNA: %S:12,51 %C:26,25 nuDNA: %S:69,55 %C:30,55	Sant Pau del Camp (Barcelona, España)	mtDNA	12	3	33,33 %	25,00 % (ambiental) 8,33 % (carry-over)	sí (investigador, antropólogo ambiental)	25,00 %
	Chaves (Huesca, España)	mtDNA	3	3	0,00 %	0,00 %	no	100,00 %
	Can Sadurní* (Barcelona, España)	mtDNA	Mínimo 7 (22 muestras)	7 (10 muestras)	4,54 %	31,81 % (investigador) 9,06 % (ambiental) 4,54 % (arqueólogo)	sí (investigador)	42,00 % (en muestras)
		nuDNA	Mínimo 5 (6 muestras)	Solo perfiles parciales y no replicables	69,55 %	30,55 % (ambiental / artefactos)	-	0 %

*Para las muestras procedentes del yacimiento de Can Sadurní no fue posible calcular los porcentajes a nivel de número de individuos porque se trataba de dientes sueltos. Por tanto se calculó a nivel de muestras.

En algunos casos, a pesar de que no se detectaran contaminaciones a nivel de secuenciación directa de los amplificadores, fue posible encontrar moléculas contaminantes puntuales durante el análisis del alineamiento de las secuencias clonadas. Sin embargo, se trataba de contaminaciones traza que no afectaron al resultado final, puesto que fue posible eliminarlas a la hora de establecer el haplotipo consenso.

En su conjunto, la eliminación de muestras por contaminación fue la predominante sobre la eliminación por falta de amplificación (en promedio incidió en el 31,56 % de las muestras analizadas, frente al 10,72 % de las muestras que no proporcionaron resultados). Ello sugiere cierta dificultad en la detección de los contaminantes en las primeras fases de análisis, como han apuntado otros autores (Malmström et al., 2005).

Este patrón se mantiene a nivel de los distintos periodos estudiados. Es interesante notar que no existen diferencias considerables entre ellos y que a este nivel, las muestras neolíticas antiguas son las que han sido descartadas en menor medida. En este caso, también podemos destacar que no hay una relación entre antigüedad y porcentajes de contaminación o ausencia de amplificación. Aquí juegan un papel importante muchos factores, como el estado de preservación de los restos y la incidencia de contaminantes de diferente origen. A este propósito, se incluyen en la **Tabla D 3** las principales fuentes de contaminación responsables de la eliminación de las muestras durante los análisis, que pueden proceder de los manipuladores, del ambiente o de otras muestras analizadas en paralelo.

La amplificación positiva de un blanco de extracción o de PCR solo permite detectar la contaminación durante el análisis en el laboratorio y no en las fases anteriores. Como ya se ha comentado, las muestras son susceptibles de resultar contaminadas por parte de arqueólogos o antropólogos que las hayan manipulado previamente a su análisis. La limpieza superficial de las muestras y su irradiación con luz ultravioleta fueron los tratamientos empleados para prevenir este tipo de contaminación. Sin embargo, los resultados obtenidos apuntan a que en ciertos casos este procedimiento no es

suficiente. Fue posible detectar contaminación por parte de arqueólogos o antropólogos, secuenciando directamente los amplificados en el 0,32 % de los individuos analizados. Sin embargo, en otros casos ésta no se detectó hasta el análisis de los alineamientos de secuencias clonadas (ver **Tabla D 3**). Por ejemplo, en el caso de una muestra procedente del yacimiento de Can Sadurní se detectó el perfil de uno de los arqueólogos mediante secuenciación directa de los amplificados. Sin embargo, en el caso de otras dos muestras del yacimiento de Sant Pau del Camp, se detectó una contaminación parcial de las secuencias clonadas que procedía del antropólogo que llevó a cabo el lavado con ácido acético ya descrito en el **Apartado D.2.1**. En el primer caso (Can Sadurní), la contaminación podría deberse al hecho de que el arqueólogo hubiera manipulado la muestra sin protección y permitiendo la penetración de su material genético en el interior de la pieza a través de alguna pequeña fisura presente en la raíz. También es posible que el ADN no hubiera penetrado en el interior del diente, quedándose depositado en la superficie sin poder ser eliminado eficazmente mediante el proceso de limpieza inicial. En el segundo caso (Sant Pau del Camp), podría tratarse de una situación parecida. Sin embargo, en este caso el empleo de ácido acético para su limpieza podría haber afectado a la degradación del ADN endógeno, facilitando la amplificación de moléculas contaminantes.

También merece la pena destacar la elevada presencia de contaminaciones en todos los conjuntos muestrales por parte del investigador responsable de los análisis, que manipuló las muestras desde su llegada al laboratorio. En total, se descartó el 12,67 % de las muestras analizadas por esta razón. También en este caso es posible que algunas de las muestras presentaran el mismo haplotipo que el investigador, pero se prefirió no considerar estos resultados como endógenos para asegurar la falta de contaminaciones. Sin embargo, en la mayoría de los casos, la secuencia del investigador se obtenía alternativamente a la ausencia de amplificación. Este dato apunta a que se tratara de muestras con escasa concentración de ADN, más vulnerables a la contaminación durante la amplificación con el *kit* empleado.

Otro tipo de contaminación detectada en un elevado porcentaje en las muestras estudiadas (16,70 %) es la contaminación ambiental. El origen de esta contaminación

no se pudo relacionar ni con los reactivos, ni con los manipuladores ni con otras muestras analizadas en paralelo. Probablemente se trataba de moléculas flotantes, presentes en el ambiente, que persistían a pesar de todas las precauciones tomadas, como son el empleo de material estéril y desechable, el análisis en laboratorios exclusivos y de acceso limitado y la limpieza constante de las superficies de trabajo. Al igual que en el caso anterior, probablemente este tipo de contaminación es más propenso a manifestarse en el caso de muestras con bajo contenido de ADN endógeno en las que se amplifican pequeñas cantidades de contaminantes ambientales por la elevada sensibilidad del *kit* empleado.

En la **Tabla D 3** también es posible ver que se detectaron contaminaciones puntuales en los alineamientos de secuencias clonadas, procedentes de los manipuladores, del ambiente o de otras muestras analizadas en paralelo. Esto indica que el control y la identificación de la contaminación son importantes en todas las fases del análisis para una correcta interpretación de los resultados. Así, para la determinación del haplotipo consenso, las secuencias contaminantes se identificaron y no se consideraron (para más detalle véanse los alineamientos de clones de la **Tabla S7** del **ARTÍCULO 3** y del **Anexo II del ARTÍCULO 4**).

D.3.2 REPLICACIÓN DE LOS ANÁLISIS EN UN SEGUNDO LABORATORIO

Además del control de la contaminación, es imprescindible la replicación de los resultados a diferentes niveles con el fin de demostrar la autenticidad de los resultados. Así para el desarrollo de la presente tesis doctoral, todos los resultados aquí incluidos se replicaron intra-laboratorio y siempre que hubiera muestra disponible también en un laboratorio diferente (LAPP-PACEA, *Université de Bordeaux 1* – *CNRS, Talence, Francia*).

La replicación de los resultados dentro del mismo laboratorio incluía la obtención de resultados replicables a partir de uno o más extractos de ADN procedentes de un mismo individuo. La replicación de los resultados inter-laboratorio se refiere a la obtención de perfiles coincidentes con los previamente obtenidos en el laboratorio de

Madrid. Las muestras correspondientes a los períodos neolítico y de la Edad de los metales (Calcolítico y Edad del Bronce) se replicaron en un segundo laboratorio (LAPP-PACEA, *Université de Bordeaux 1 – CNRS, Talence, Francia*). Se decidió replicar únicamente las muestras más antiguas de acuerdo a lo descrito en Pääbo y colaboradores (2004). En el caso de que se dispusiera de diferentes piezas óseas o dentales, se repitieron todos los análisis desde la limpieza de la muestra hasta su trituración, extracción y amplificación. En otros casos, por falta de disponibilidad de muestras, se repitieron los análisis a partir de una fracción de polvo de hueso o diente obtenida en el laboratorio de origen.

El porcentaje de éxito de las diferentes replicaciones realizadas y su resultado se detallan en la **Tabla D 4**. Aquí es posible destacar un porcentaje de éxito en la obtención de resultados replicables en los dos laboratorios relativamente alto (62,96 %), lo que refuerza la validez de la metodología empleada. Además, indica que los estándares empleados en ambos laboratorios permiten la obtención de resultados válidos, lo que añade validez también a los restantes resultados en los que no fue posible realizar un segundo análisis por falta de disponibilidad de muestras. Además, para el restante 47,04 % de resultados no replicables es posible proporcionar una explicación relacionada con problemáticas concretas del laboratorio LAPP-PACEA (*Université de Bordeaux 1 – CNRS, Talence, Francia*). Aquí se detectaron diferentes fenómenos de contaminación e inhibición en algunas de las muestras, que no se habían manifestado en el laboratorio de Madrid.

Antes de implementar el protocolo S-B también en este laboratorio, se utilizó el método empleado comúnmente en este laboratorio para la extracción de las primeras muestras. Se constató que este método, que emplea fenol y cloroformo (Deguilloux et al., 2011 b), no conseguía eliminar eficazmente las moléculas inhibidoras de la PCR. Las muestras que se extrajeron según este protocolo no proporcionaron resultados. Este es el caso de la muestra 25SP02 del yacimiento de Sant Pau (ver **Tabla D 4**). Probablemente el fallo de este protocolo esté relacionado con una eliminación ineficiente de la porción fenólica, que se comporta como un potente inhibidor de la PCR (Wilson, 1997).

Tabla D 4 Relación del número de muestras analizadas en un segundo laboratorio. Se detallan los porcentajes de replicación exitosos así como las causas de la ausencia de replicación.

Período	Yacimiento	Marcadores analizados	Nº muestras analizadas 2º laboratorio	% replicación resultados	Posibles causas ausencia replicación	Comentario
Edad de los Metales (57,14 %)	Mas d'Abad (Castellón, España)	mtDNA	1	100,00 %	-	-
	Sepulcre Segond de la Costa (Castellón, España)	mtDNA	1	100,00 %	-	-
	Torre d'en Domenèch (Castellón, España)	mtDNA	1	100,00 %	-	Obtención información genética adicional.
	Cova de Malpaso (Castellón, España)	mtDNA	3	0,00 %	Contaminación personal laboratorio	-
	Racó de Raca (Castellón, España)	mtDNA	1	0,00 %	Contaminación cebadores	-
	Assud de Villarreal (Castellón, España)	mtDNA	1	100,00 %	-	-
Neolítico Antiguo (42,85 %)	Sant Pau del Camp (Barcelona, España)	mtDNA	1	0,00 %	No amplifica	-
	Chaves (Huesca, España)	mtDNA	3	66,66 %	No amplifica (método de extracción)	-
	Can Sadurní (Barcelona, España)	mtDNA	3	100,00 %	-	Replicación de un solo fragmento en el caso de dos muestras.

Por otro lado, a pesar de que los análisis fueron llevado a cabo en ambos laboratorios por el mismo investigador (la autora de la presente tesis doctoral) se detectaron contaminaciones del otro investigador que estaba usando las instalaciones en ese mismo período en tres muestras procedentes de la Cova de Malpaso (ver **Tabla D 4**). Este fenómeno no se había puesto de manifiesto en el laboratorio de Madrid, probablemente debido al hecho de que en este laboratorio cada investigador empleaba material diferente (como, por ejemplo, los juegos de pipetas), mientras que en el laboratorio de Talence no fue posible separar este material debido a la falta de disponibilidad del mismo. Además, en el laboratorio de Talence, se llevaron a cabo diferentes extracciones en presencia de otro investigador, mientras que en Madrid el acceso al laboratorio estaba restringido a un solo investigador durante la fase de extracción de ADN. Adicionalmente, en el segundo laboratorio se detectaron contaminaciones puntuales procedentes de los cebadores (Deguilloux et al., 2011 a),

como en el caso de la muestra RR3 del yacimiento de Racó de Raca (ver **Tabla D 4**). Estas contaminaciones nunca fueron detectadas en el laboratorio de origen. Todos estos datos, podrían explicar la falta de replicación del 47,04 % de los resultados.

A pesar de estas dificultades, fue posible obtener resultados replicados para la mayoría de las muestras analizadas. Además en ciertos casos fue posible obtener información genética complementaria a la previamente obtenida en el laboratorio de origen, completando haplotipos parciales (ver **Tabla D 4**).

A la luz de los resultados obtenidos en la presente tesis doctoral, es posible afirmar que la replicación en un segundo laboratorio representa un criterio de autenticidad muy importante en el caso de muestras por las cuales resulta difícil establecer un haplotipo consenso; bien por la presencia de un elevado daño molecular o bien por la presencia de contaminantes de origen desconocido. En estos casos, la replicación inter-laboratorio puede permitir la identificación del haplotipo consenso, idéntico para dos muestras analizadas en diferentes laboratorios, así como la detección de contaminantes ambientales presentes en un laboratorio concreto. Sin embargo, también se ha demostrado que en el caso de muestras no problemáticas es posible obtener resultados replicables no contaminados. Así para muestras recientes y muy bien preservadas, como las medievales de la Iglesia de San Esteban de Cuéllar (**ARTÍCULO 5**), se consideró que la replicación de resultados de nuDNA intra-laboratorio pudiera representar un criterio suficiente para su validación, puesto que no se detectó ningún tipo de contaminación y se incluía una amplia replicación de resultados a todos los niveles (análisis de diferentes muestras de un mismo individuo y amplificaciones independientes de un mismo extracto).

D.3.3 CONSIDERACIONES GENERALES SOBRE LA VALIDACIÓN DE LOS RESULTADOS

Todos los criterios de criba e interpretación de los resultados descritos en los apartados de esta sección permitieron obtener resultados validados a partir de un conjunto de 32 individuos antiguos de épocas históricas y prehistóricas.

En el caso de las muestras en las que se analizó el mtDNA se confirmó el haplogrupo correspondiente mediante la amplificación de SNPs en la región codificante y en el caso de las muestras neolíticas se calculó la incidencia del daño molecular (ver **Material y Métodos del ARTÍCULO 3**). Gracias a estos análisis fue posible confirmar los resultados obtenidos, puesto que existía concordancia filogenética entre los haplotipos y los haplogrupos determinados mediante el análisis de SNPs, localizados en la región codificante. Además, el análisis del daño molecular confirmó la autenticidad de los resultados, dado que se detectó la prevalencia de transiciones de Tipo II (CG→TA, 68 %), como apuntan otros autores (Brotherton et al., 2007; Gilbert et al., 2007).

Las consideraciones sobre la contaminación y la replicación de los resultados aquí descritos ponen de manifiesto la gran importancia en la obtención de todos los perfiles de los manipuladores para poder descartar su perfil del consenso. También juega un papel importante el rastreo de las tandas de extracción y amplificación de las muestras, para poder identificar fenómenos de *carry-over*. Sin embargo, la contaminación ambiental representa el tipo de contaminante más difícil de determinar ya que se desconoce su origen. Los resultados obtenidos en la presente tesis doctoral apuntan a que se trata de moléculas flotantes presentes en el ambiente de trabajo, pues este tipo de contaminación no es replicable. Es a este nivel que la replicación de los resultados inter-laboratorio juega un papel importante, permitiendo identificar el haplotipo consenso y detectar las contaminaciones ambientales de un laboratorio concreto.

DISCUSIÓN DE LOS OBJETIVOS ESPECÍFICOS – ESTUDIO DIACRÓNICO-POBLACIONAL

D.4 Estimación de la continuidad/discontinuidad genética entre los pobladores prehistóricos (neolíticos, calcolíticos, de la Edad del Bronce e íberos) y actuales del noreste de la Península Ibérica (Comunidad Valenciana, Cataluña y Aragón) y evaluación del impacto genético del Neolítico en esta misma área.

La continuidad o discontinuidad genética entre los pobladores del noreste de la Península Ibérica ha sido evaluada desde diferentes perspectivas a lo largo del desarrollo de la presente tesis doctoral. Por un lado se ha analizado la composición de haplotipos y haplogrupos mitocondriales de las diferentes poblaciones estudiadas. Por otro lado se han cuantificado las aportaciones de los distintos pobladores mediante métodos estadísticos y computacionales.

D.4.1 COMPARACIÓN DE LA COMPOSICIÓN DE HAPLOTIPOS Y HAPLOGRUPOS DE LAS POBLACIONES NEOLÍTICAS, CALCOLÍTICAS, DE LA EDAD DEL BRONCE E ÍBERAS DEL NORESTE PENINSULAR ENTRE SÍ Y CON OTRAS POBLACIONES, ANTIGUAS Y ACTUALES, DE LA MISMA REGIÓN (ARTÍCULOS 3 Y 4)

Para la interpretación de los resultados en un contexto diacrónico, se compararon los haplotipos y los haplogrupos obtenidos en los diferentes periodos estudiados entre sí y también con aquellos descritos en la literatura y procedentes de la misma región, el noreste peninsular.

Puesto que los resultados aquí discutidos se analizaron separadamente en dos artículos incluidos en la presente tesis doctoral (**ARTÍCULOS 3 y 4**) se vuelven a presentar todos los resultados en la **Tabla D 5** de este apartado a fin de llevar a cabo una valoración conjunta de los mismos. Además, se presentan en la **Tabla D 6** todos los haplotipos compartidos entre las poblaciones estudiadas y las poblaciones antiguas

disponibles en la literatura con las que se han comparado los resultados. En la **Figura D 2** también se muestran las frecuencias de haplogrupos con fines comparativos. Se incluyen aquellas que proceden de la misma región geográfica correspondientes al Neolítico Medio (Sampietro et al., 2007) y pertenecientes al período íbero (Sampietro et al., 2005), así como otras poblaciones mesolíticas y neolíticas centroeuropeas (Haak et al., 2005, 2010; Bramanti et al., 2009), del norte de Europa (Malmström et al., 2009) y del Sur de Francia (Lacan et al., 2011; Deguilloux et al., 2011 b).

Se agruparon las muestras estudiadas en tres períodos (Neolítico Antiguo, Edad de los Metales e íberos) para que cada grupo alcanzara un número mínimo de 4 individuos. Por otra parte, a pesar de que la cronología de los íberos se incluya dentro de la Edad de los Metales (más concretamente en la Edad del Hierro), se consideraron como un grupo poblacional diferente debido a su reciente cronología y a otros factores relacionados con los rituales de enterramiento. Las costumbres funerarias íberas se caracterizaban por la cremación de los cuerpos. Las muestras analizadas pertenecen a aquella pequeña fracción de población que tuvo un tratamiento diferente, puesto que para un análisis exitoso de ADN antiguo es necesario emplear huesos y dientes no quemados. Más concretamente, las muestras procedentes de Los Villares y de Tossal de Sant Miquel se corresponden con individuos infantiles, mientras que las de Puig de la Nau son individuos adultos. En el caso de las inhumaciones infantiles, se ha sugerido en la literatura que se podría tratar de ritos propiciatorios o de protección de la ciudad o de la familia, que podrían incluso implicar el sacrificio humano (Oliver Foix, 2004; Barrial Jove, 1990). La localización de restos animales en los mismos lugares de enterramientos infantiles apoyaría su carácter sacrificial (Oliver Foix, 2004).

En lo que concierne a los individuos adultos no incinerados, existen varias hipótesis alternativas. Es posible que se tratara de un sacrificio, como en el caso de los individuos infantiles, o de una “mala muerte”, expresión acuñada en los años 90 (Dedet, 1992) y que hace referencia a aquellas muertes acontecidas de manera inusual para la sociedad. En el caso concreto del yacimiento de Puig de la Nau, la presencia de individuos adultos no incinerados está asociada a muertes violentas, más concretamente al culto de las cabezas cortadas (ritual de origen celta que consiste en

cortar la cabeza del enemigo y exponerla como trofeo) y a la exposición pública de los cuerpos (Oliver Foix, 2004). Todo ello podría indicar que se trataba de una población foránea enemiga.

Por otro lado, entre Calcolítico y Edad del Bronce no hay referencias arqueológicas que sugieran movimientos poblacionales importantes ni de ritos funerarios diferenciales y por esta razón ambos períodos se incluyeron en el mismo grupo poblacional: la Edad de los Metales. En lo que concierne al Neolítico Antiguo se agruparon las poblaciones Cardiales y Post-Cardiales, puesto que ambas están ligadas a las primeras fases de desarrollo del Neolítico en la Península. En el caso concreto del yacimiento epicardial estudiado aquí (San Pau del Camp) también se registraron niveles cardiales en el mismo yacimiento, pero carentes de restos humanos.

Tabla D 5 Tabla resumen de los resultados obtenidos en los **ARTÍCULOS 3 y 4** (respectivamente **Tabla A1** y **Tabla 1**). Los haplotipos coincidentes están marcados con el mismo color (dentro de esta tabla y con la **Tabla D 6**).

	Yacimiento	Período	Datación	Individuo	Haplotipo	SNPs tipados	Haplogrupo
Íberos	Tossal Sant Miquel	Edad del Hierro	Siglo IV-III a.C.	TSM2	16311C	10550G	K
	Los Villares	Edad del Hierro	Siglo IV-III a.C.	LV1	16298C	4580G	HV0
	Puig de la Nau	Edad del Hierro	Siglo V a.C.	PB14	16293T 16298C	4580G	HV0
	Puig de la Nau	Edad del Hierro	Siglo V a.C.	PB20	16293T 16298C	4580G	HV0
Edad de los Metales	Maçanet	Edad del Bronce	1.500-1000 a.C.	1MN	16069T 16126C	13708A	J
	Mas d'Abad	Edad del Bronce	1.500-1000 a.C.	3MA	16189C 16224C 16311C	10550G	K
	Sepulcre Segond de la Costa	Calcolítico	2500-2000 a.C.	COST1	16136C 16192T 16270T 16304C	3197C 7028T 10398A 10873T 10400C	U5b3
	Sepulcre Segond de la Costa	Calcolítico	2500-2000 a.C.	COST4	CRS	4580G 7028T 14766C	HV
	Torre d'en Domenèch	Calcolítico	2500-2000 a.C.	1TD	16224C 16311C	10550G	K
	Cova de Malpaso	Calcolítico	2500-2000 a.C.	1MP	16223T 16362C	10873T 10400C 14766T	N*
	Racó de Raca	Calcolítico	2500-2000 a.C.	RR1	16189C 16223T 16278T	6371T 10873T 14766T	X2
	Assud de Villarreal	Calcolítico	2500-2000 a.C.	1AV	16298C	4580A	V
Neolítico Antiguo	Sant Pau del Camp	Neolítico Epicardial	4.250-3.700 a.C.	6SP0102	16224C 16311C	10550G	K
	Sant Pau del Camp	Neolítico Epicardial	4.250-3.700 a.C.	26SP0102	16218T 16328A (16362C)	7028C	H20
	Sant Pau del Camp	Neolítico Epicardial	4.250-3.700 a.C.	27SP0102	16147T 16223T 16362C	7028T 10238T 10873T	N*
	Chaves	Neolítico Cardial	5.329-4.999 a.C.	1CH0102	16224C 16311C	10550G	K
	Chaves	Neolítico Cardial	5.329-4.999 a.C.	2CH0102	CRS	7028C	H
	Chaves	Neolítico Cardial	5.329-4.999 a.C.	3CH01	16129A	7028C	H
	Can Sadurní	Neolítico Cardial	5.475-5.305 a.C.	CSA0511	16223T 16362C	4833A 5178C 10238T 10398A 10400C 10873T	N*
	Can Sadurní	Neolítico Cardial	5.475-5.305 a.C.	CSA09	16223T 16362C	4833A 5178C 10238T 10398A 10400C 10873T	N*
	Can Sadurní	Neolítico Cardial	5.475-5.305 a.C.	CSA152223	16224C 16311C	10550G	K
	Can Sadurní	Neolítico Cardial	5.475-5.305 a.C.	CSA16	16362C	7028C	H
	Can Sadurní	Neolítico Cardial	5.475-5.305 a.C.	CSA24	16136C 16192T 16270T	3197C 14766T	U5*
	Can Sadurní	Neolítico Cardial	5.475-5.305 a.C.	CSA26	16183C 16189C(het) 16223T 16278T	1719G 6371GT 7028T	X1
	Can Sadurní	Neolítico Cardial	5.475-5.305 a.C.	CSA29	16147T 16223T 16362C	10238T 10398A 10400C 10873T	N*

Tabla D 6 Haplotipos compartidos entre las muestras estudiadas y las disponibles en la literatura. Se marcan las coincidencias de haplotipos con los mismos colores empleados en la **Tabla D 5**.

Haplotipo	Yacimiento	Localización	Época	Frecuencia haplotípica	Referencia
16298C	Treilles	Francia	Neolítico Final	10,34 % (3/29)	Lacan et al., 2011
	Derenburg, Halberstadt	Alemania	Neolítico LKB	4,76 % (2/42)	Haak et al., 2010
16069T 16126C	Treilles	Francia	Neolítico Final	20,69 % (6/29)	Lacan et al., 2011
	Camí de Can Grau	España	Neolítico Medio	18,18 % (2/11)	Sampietro et al., 2007
	Derenburg, Seehausen	Alemania	Neolítico LKB	7,14 % (3/42)	Haak et al., 2005 y 2010
	Ostorf	Alemania	Mesolítico (3000 a.C.)	4,54 % (1/22)	Bramanti et al., 2009
16224C 16311C	Treilles	Francia	Neolítico Final	6,89 % (2/29)	Lacan et al., 2011
	Derenburg, Flomborn	Alemania	Neolítico LKB	7,14 % (3/42)	Haak et al., 2005 y 2010
	Ostorf	Alemania	Mesolítico (3200 a.C.)	4,54 % (1/22)	Bramanti et al., 2009
	Illa d'en Reixac	España	Íbero	5,88 % (1/17)	Sampietro et al., 2005
CRS	Frälsegården	Suecia	Neolítico TRB (3.500–2.500 a.C.)	33,33 % (1/3)	Malmström et al., 2009
	Hohler Fels	Alemania	Magdaleniense (13.400 a.C.)	4,54 % (1/22)	Bramanti et al., 2009
	Illa d'en Reixac	España	Íbero	5,88 % (1/17)	Sampietro et al., 2005
	Treilles	Francia	Neolítico Final	20,69 % (6/29)	Lacan et al., 2011
	Camí de Can Grau	España	Neolítico Medio	27,27 % (3/11)	Sampietro et al., 2007
	Derenburg, Eilsleben, Asparn Schletz	Alemania	Neolítico LKB	9,52 % (4/42)	Haak et al., 2005 y 2010
16183C 16189C 16223T 16278T	Treilles	Francia	Neolítico Final	10,34 % (3/29)	Lacan et al., 2011
16311C	Derenburg, Flomborn	Alemania	Neolítico LKB	9,52 % (4/42)	Haak et al., 2005 y 2010
	Ajvide	Mar Báltico	Mesolítico PWC (2.800–2.000 a.C.)	10,52 % (2/19)	Malmström et al., 2009

D.4.1.1 COMPOSICIÓN HAPLOTÍPICA DE LAS POBLACIONES

En el conjunto de muestras analizadas destaca la elevada diversidad haplotípica: sobre un total de 25 especímenes analizados y validados, se encontraron 15 haplotipos

únicos (79 %) y 4 compartidos (21 %). En función de la distribución de los haplotipos compartidos se realizaron inferencias a nivel familiar y poblacional.

Por ejemplo, el haplotipo 16224C 16311C resulta compartido por dos individuos neolíticos cardiales de dos yacimientos diferentes (1CH0102 y CSA152223), un espécimen neolítico epicardial (6SP0102) y otro calcolítico (1TD). Esto indica que se trata de un tipo mitocondrial que estaba presente en el Península Ibérica al menos desde el comienzo del Neolítico. También se ha hallado en otra muestra íbera de la misma región (5,88 %) (Sampietro et al., 2005), en dos muestras neolíticas finales francesas (6,89 %) (Lacan et al., 2011), así como en el 7,14 % de muestras neolíticas (Haak et al., 2005, 2010) y en una muestra mesolítica (4,54 %) de datación reciente (3.200 a.C.) (Bramanti et al., 2009) de Europa central (ver **Tabla D 6**).

Teniendo en cuenta que se ha detectado en diferentes periodos a pesar del reducido número de muestras, esto podría indicar cierta frecuencia y continuidad de este haplotipo en la Península Ibérica y en Europa. Esta continuidad se ve confirmada por su presencia en la población actual de la región estudiada, así como en Europa, con frecuencias del 6 % (García et al., 2011) y 3 % (Richards et al., 2000) respectivamente.

Dos individuos íberos (PB14 y PB20) podrían estar emparentados a lo largo de la línea materna puesto que comparten el haplotipo 16293T 16298C y proceden del mismo yacimiento. Para inferir probabilidades de relaciones familiares es necesario conocer la composición genética de la población de origen y más concretamente la frecuencia del haplotipo en cuestión. Sin embargo, el hecho de que no haya sido posible detectar este haplotipo específico en las demás poblaciones antiguas comparadas (ver **Tabla D 6**), en la población europea actual (Richards et al., 2000; Parson et al., 2004; Lee et al., 2008) y en el noreste de la Península Ibérica (García et al., 2011), incrementa la probabilidad de una relación directa entre los dos individuos a lo largo de la línea materna.

Una situación parecida afecta a las muestras CSA09 y CSA152223 del yacimiento de Can Sadurní, que comparten el haplotipo 16223T 16362C. En este caso, además, el

análisis antropológico (Blasco et al., 2005; Cabellos, 2008) indica que se trata de un individuo infantil y otro adulto respectivamente. Como en el caso precedente, se trata de una variante mitocondrial no representada en la población antigua y actual de la región de procedencia y tampoco a nivel europeo, apuntando a la posibilidad de que se trate de dos individuos relacionados a lo largo de la línea materna.

Otro haplotipo no representado en las poblaciones actuales es el 16147T 16223T 16362C, que es compartido por los individuos 27SP0102 (Sant Pau del Camp) y CSA29 (Can Sadurní). Como ocurre para los dos casos anteriores, tampoco se encuentra representado en las poblaciones modernas y antiguas con las que se ha comparado. Estos datos apuntan a la presencia de continuidad genética entre el Neolítico Cardial y el sucesivo Epicardial de la misma región geográfica. Además, como ya se ha comentado en el **Apartado D.4.1**, el yacimiento de Sant Pau del Camp presenta niveles cardiales, carentes de restos humanos. Alternativamente, estos resultados podrían indicar que ambas poblaciones compartieran un mismo sustrato poblacional. Se trata de dos yacimientos separados por una distancia de 20 Km, mucho más cercanos entre sí que al yacimiento neolítico cardial de Chaves, localizado aproximadamente a 200 Km de ambos.

D.4.1.2 COMPOSICIÓN DE HAPLOGRUPOS DE LAS POBLACIONES

Pasando al análisis de haplogrupos, se observa como el total de los 19 haplotipos se asigna a 9 haplogrupos (ver **Tabla D 5**). En este caso también merece la pena analizar aquellos compartidos (4) y únicos (5) en una perspectiva filogenética.

El haplogrupo más representado es el K, que se encuentra en todos los grupos poblacionales y períodos analizados, puesto que lo comparten 2 individuos neolíticos cardiales, 1 epicardial, 1 calcolítico, 1 de la Edad del Bronce y otro íbero. Como ya se apuntó para el haplotipo 16224C 16311C, que es el haplotipo raíz del haplogrupo K y el más frecuente, se podría tratar de un marcador de continuidad genética entre los períodos estudiados. Se trata de un haplogrupo típicamente europeo, que presenta una frecuencia del 9 % en la población actual (Richards et al., 2000).

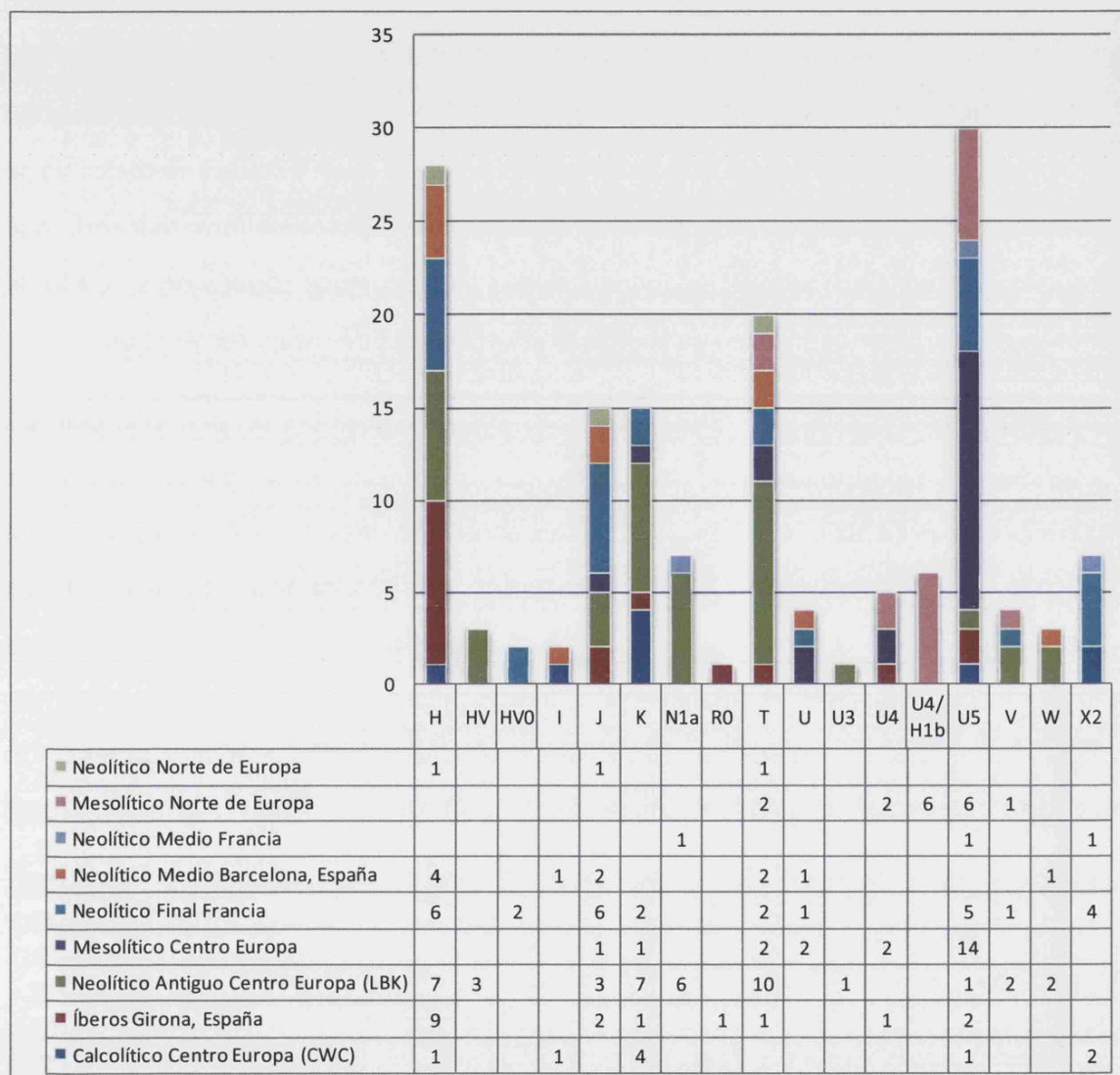
Además, este haplogrupo presenta una elevada frecuencia en muestras neolíticas acerámicas de Próximo Oriente, previamente estudiadas por nuestro grupo de investigación (Fernández, 2005). En este sentido, la elevada frecuencia del haplogrupo K en estas regiones podría estar revelando una conexión demográfica entre ellas.

El haplogrupo K se ha detectado en una sola muestra mesolítica de Alemania (Bramanti et al., 2009), representando el 4,54 % de la población estudiada (ver **Figura D 2**). Sin embargo, esta muestra está asociada a una datación muy reciente (3.200 a.C.) (Bramanti et al., 2009). Aunque este haplogrupo estuviera presente en los pobladores mesolíticos europeos antes de la llegada del Neolítico, la elevada frecuencia del K en las muestras analizadas podría apuntar a un incremento de su frecuencia debido a la llegada de nuevos pobladores. Su elevada frecuencia en muestras neolíticas de Europa Central (~12 %) resulta compatible con esta hipótesis (Haak et al., 2005, 2010). Este haplogrupo se encuentra también representado en períodos posteriores, siendo su frecuencia del 7 % en la población neolítica final de Treilles (Francia) (Lacan et al., 2011) y del 44 % en el conjunto de enterramientos múltiples de la cultura calcolítica de la cerámica cordada de Eulau, Alemania (Haak et al., 2008) (ver **Figura D 2**).

Sin embargo, desde la genética de poblaciones actuales, no se ha sugerido una conexión del haplogrupo K con la expansión neolítica, sino que se han propuesto como marcadores neolíticos los haplogrupos J, T1 y U3, debido a la asociación entre este período y su datación molecular (Richards et al., 2000). Como se puede observar en el caso de las muestras analizadas en la presente tesis doctoral, ninguno de ellos está representado en las muestras neolíticas. Solamente se detectó un haplogrupo J en una muestra de la Edad del Bronce (1MN del yacimiento de Maçanet, ver **Tabla D 5**), varios miles de años posterior.

Entre las otras muestras neolíticas analizadas por otros autores, es posible encontrar un solo representante del haplogrupo U3 (2 %) en Europa central (Haak et al., 2005, 2010), y ningún T1 (ver **Figura D 2**).

Figura D 2 Gráfico de la distribución de haplogrupos en las poblaciones antiguas con las que se han comparado los datos de la presente tesis. Referencias bibliográficas: Calcolítico Centro Europa (CWC) [Haak et al. 2008], íberos Girona, España [Sampietro et al. 2005], Neolítico Antiguo Centro Europa (LBK) [Haak et al. 2005, 2010], Mesolítico Centro Europa [Bramanti et al. 2009], Neolítico Final Francia [Lacan et al. 2011], Neolítico Medio Barcelona, España [Sampietro et al. 2007], Neolítico Medio Francia [Deguilloux et al. 2011], Mesolítico Norte de Europa [Malmström et al. 2009], Neolítico Norte de Europa [Malmström et al. 2009].



En cambio, el haplogrupo J se encuentra en un 18 % de las muestras del Neolítico Medio de la provincia de Barcelona (Sampietro et al., 2007), en el 7 % de las muestras del Neolítico LBK de Europa central (Haak et al., 2005, 2010), en una de las tres muestras neolíticas del norte de Europa (Malmström et al., 2009) y en el 20 % de muestras del Neolítico Final del yacimiento francés de Treilles (Lacan et al., 2011) (ver Figura D 2).

Los haplogrupos T1 y U3 tampoco están representados en las poblaciones mesolíticas del norte y centro Europa. Mientras que el haplogrupo J está presente en el 5 % de estas poblaciones (un individuo de Europa central y uno de Europa septentrional, ver **Tabla D 6**) (Bramanti et al., 2009; Malmström et al., 2009) (ver **Figura D 2**). Estos resultados sugieren que el haplogrupo J quizás no estuviera necesariamente asociado a la llegada de nueva población neolítica. Hay que tener en cuenta, sin embargo, que algunos de los individuos de cultura mesolítica citados están asociados a una datación muy reciente (alrededor de 3.000 a.C.), momento en el cual la cultura neolítica ya se había expandido prácticamente por toda Europa. Una hipótesis alternativa sería que en este caso concreto se tratase de inmigrantes neolíticos que cambiaron su estilo de vida para adaptarse al entorno volviendo a una economía de caza y recolección.

Sin embargo, a la luz de estos resultados cabe la posibilidad de replantear las inferencias filogenéticas basadas en la genética de poblaciones actuales acerca de las contribuciones neolíticas al acervo europeo actual. En primer lugar, como ya se ha comentado en el **Apartado I.4.1.2**, la asociación entre la datación molecular de un haplogrupo no tiene por qué coincidir con su expansión. En segundo lugar, las dataciones moleculares suelen estar asociadas a amplios rangos de error estándar (ver **Figura I 11**). Además, no hay acuerdo en la comunidad científica acerca de la precisión de los métodos de datación molecular del mtDNA (Bandelt, 2007; Ho et al., 2007; Howell et al., 2008), puesto que no es posible realizar una calibración adecuada (Pulquério y Nichols, 2007) y existen discrepancias entre los datos estimados a partir de *pedigrees* y los de análisis filogenéticos (Howell et al., 2008).

El segundo haplogrupo más representado en las poblaciones estudiadas es el N*, que caracteriza principalmente las poblaciones del Neolítico Cardial y Epicardial, pero también se detectó en un individuo de época calcolítica. Se trata de un haplogrupo muy particular que actualmente se encuentra presente en Europa Oriental (población Serbia, 1,71 %) y en poblaciones de Próximo Oriente (2 % de media) así como en población judía askhenazi europea (0,32 %) y de Próximo Oriente (1,68 %) (para más detalles y referencias bibliográficas ver **Tabla S3** del **ARTÍCULO 3**), pero que no está representado en ninguna otra población europea. Sin embargo, un espécimen de Cro-

Magnon de hace 25.000 años localizado en el Sur de Italia fue clasificado como perteneciente a este haplogrupo, aunque el elevado grado de degradación del ADN no permitió la amplificación de todos los SNPs que caracterizan al haplogrupo N* (Caramelli et al., 2003). El haplotipo de esta muestra es, sin embargo, diferente al encontrado en las muestras 1MP, 27SP0102, CSA0511, CSA09 y CSA29.

El caso del haplogrupo N* es muy similar al descrito para muestras neolíticas centroeuropeas (Haak et al., 2005), donde también se encontró una frecuencia muy elevada de un haplogrupo actualmente raro en Europa (0.2 %), el N1a (ver **Figura D 2**). Ambos hallazgos podrían indicar que las primeras poblaciones neolíticas presentaban unos linajes característicos que sucesivamente se perdieron por deriva genética. El hecho de que el N* caracterice a los habitantes del noreste peninsular desde el Neolítico antiguo Cardial hasta el Calcolítico, apunta nuevamente a la presencia de cierta continuidad genética entre los diferentes pobladores, al menos a partir de la llegada de los primeros agricultores. Sin embargo, su sucesiva desaparición podría explicarse mediante mecanismos de deriva genética debido a la reducida contribución genética de la población neolítica a los habitantes autóctonos, deducida a partir de los estudios de simulación (ver **Apartado D.4.2.1**).

También el haplogrupo X1, presente en una de las muestras neolíticas cardiales analizadas (CSA26), no se encuentra en muestras europeas actuales pero sí en poblaciones de Próximo Oriente, como sucede con el N* (ver **Tabla S3 del ARTÍCULO 3**). Existe la posibilidad de que este haplogrupo, junto con el N* y quizás el K, fuera introducido en la Península Ibérica al comienzo del Neolítico por parte de poblaciones próximo-orientales. Similitudes encontradas entre los rituales de enterramiento del Neolítico de Próximo Oriente y del Neolítico Cardial peninsular (Hodder, 2007; Utrilla et al., 2008) apoyan la existencia de una relación entre ambas regiones a inicios del Neolítico. Concretamente, en el caso del yacimiento de Chaves, el individuo 1CH0102 presentaba un ritual de enterramiento típicamente próximo-oriental, en posición fetal al fondo de la fosa, con brazos y piernas flexionados y pegados contra el cuerpo. Dicho individuo pertenece al haplogrupo K, lo que viene a reforzar la hipótesis anteriormente

planteada de que este pudiera tratarse de un haplogrupo relacionado con las expansiones Neolíticas desde Próximo Oriente.

Otro de los haplogrupos representado en más de un individuo es el H, que se detectó en 4 especímenes (16 %), todos neolíticos (3 cardiales y uno post-cardial). Se trata del haplogrupo europeo más frecuente en la actualidad, con una frecuencia de alrededor del 50 % en Europa (Richards et al., 2000) y del 43 % en la población actual del noreste ibérico (García et al., 2011). Ya que actualmente se encuentra altamente representado, es difícil realizar cualquier tipo de inferencia basada en la distribución del mismo. Pero resulta interesante destacar su baja frecuencia en el conjunto de muestras estudiadas (16 %) y su ausencia en los períodos posteriores al Neolítico. Es posible que este haplogrupo no haya sido detectado en dichos períodos porque se estudió un reducido tamaño muestral. En las muestras íberas y neolíticas estudiadas por Sampietro y colaboradores (Sampietro et al., 2005, 2007) el haplogrupo H es el haplogrupo más representado (52 % y 36 % respectivamente), pero en muestras neolíticas francesas y centroeuropeas no supera el 20 % (Haak et al., 2005, 2010; Lacan et al., 2011). En conjunto, estos resultados apuntan a que la frecuencia de este haplogrupo en el pasado no habría sido tan elevada en todas las regiones europeas como lo es en la actualidad, y a que ésta habría incrementado sucesivamente probablemente debido a fenómenos de deriva genética.

Por otra parte, fue posible definir de manera más específica el sub-haplogrupo de la muestra 26SP0102 gracias a las mutaciones características de su haplotipo (Roostalu et al., 2007). En base a éstas, este haplogrupo fue clasificado como H20. Como en el caso de los haplogrupos N* y X1, este haplogrupo presenta una frecuencia más elevada en Próximo Oriente (0,56 %) que en Europa (0,18 %) (Richards et al., 2000) y representa el 2,5 % de los sub-haplogrupos de Próximo Oriente (Roostalu et al., 2007), con picos de frecuencia en Jordania, Georgia y Siria (Roostalu et al., 2007). Estos datos apoyan nuevamente el flujo génico entre los dos extremos del Mediterráneo.

Además del X1, también se pudo identificar un haplogrupo X2 en una muestra calcolítica (RR1). Al contrario del X1, el X2 está representado, aunque en bajas frecuencias, en todos los continentes. El 97,2 % de los haplogrupos X de Próximo

Oriente, del Cáucaso y de Europa pertenecen al sub-haplogrupo X2 así como el 100 % de los procedentes de Asia Central y Siberia y el 36,8 % de los norteafricanos (Reidla et al., 2003). De acuerdo con la filogeografía actual de este haplogrupo, se supone que se originó alrededor o posteriormente al último máximo glacial (Reidla et al., 2003). De acuerdo con este estudio, el haplogrupo X2 estaría representado en el sustrato paleolítico de la región. Sin embargo, es importante destacar que hasta la fecha no ha sido encontrado en ninguna muestra pre-Neolítica (Bramanti et al., 2009; Malmström et al., 2009). Sin embargo, sí ha sido detectado en varias muestras neolíticas francesas, una del yacimiento de Prissé-la-Charrière (33 %) (Deguilloux et al., 2011 b) y en 4 de de Treilles (14 %) (Lacan et al., 2011). A pesar de la antigua datación molecular de este haplogrupo (17.900 ± 2.900), en base a los datos disponibles, se podría asociar su llegada a Europa durante el Neolítico, poniendo nuevamente en discusión las inferencias basadas en la genética actual. La presencia de este haplogrupo en una muestra Calcolítica podría apuntar a una relación entre las regiones francesas y el noreste de la Península Ibérica durante el período comprendido entre el Neolítico y el Calcolítico, que podría estar relacionada con la expansión de la cultura del vaso campaniforme.

Otro haplogrupo compartido entre diferentes individuos es el HV, que incluye el sub-haplogrupo HV0. El primero caracteriza a un espécimen calcolítico (COST4), mientras que el segundo está presente en 3 de los 4 individuos íberos tipados (LV1, PB14, PB20). El HV0 se encuentra a solo dos pasos mutacionales del HV: 72C en HVR-II y 16298C en HVR-I. También el haplogrupo V, detectado en una muestra calcolítica (1AV) se encuentra a dos pasos mutacionales (15904T y 4580A) del HV0.

Desde la perspectiva de la genética de poblaciones actuales, estos haplogrupos se han asociado tradicionalmente con el sustrato paleolítico de la diversidad mitocondrial europea (Richards et al., 2000; Torroni et al., 2001). Además, se clasificó otro espécimen Cro-Magnon como perteneciente a la rama del haplogrupo HV o Pre-HV, confirmando la antigüedad del macro-haplogrupo (Caramelli et al., 2003).

En particular, se supone que el haplogrupo V se originó durante el último máximo glacial en el refugio paleolítico localizado en la franja cantábrica (Torroni et al., 1998).

La frecuencia media del haplogrupo V en la población europea actual es de un 4,8 % (Torroni et al., 1998), sin embargo este haplogrupo presenta picos máximos de frecuencia en la franja cantábrica y el País Vasco (12 %) (Torroni et al., 2001) y en el grupo Saami de Escandinavia (46,4 %) (Tambets et al., 2004).

El haplogrupo HV0 es poco frecuente en Europa, aunque tiene mayor representación en las costas mediterráneas, especialmente norteafricanas e ibéricas (Torroni et al., 2001). Teniendo en cuenta que se trata del haplogrupo del que se originó el V, desde la genética de poblaciones actuales se ha sugerido que el HV0, anteriormente llamado pre*V, se originó en Europa Oriental y se expandió hacia el oeste durante el Gravetiense (Torroni et al., 2001).

Un aspecto que resulta interesante destacar es que los haplogrupos HV0 y V no habían sido detectados hasta el momento en muestras de épocas calcolíticas e ibéricas de la Península (Izagirre y de la Rúa, 1999; Sampietro et al., 2005). Todas las muestras de época ibérica analizadas en el trabajo de Sampietro y colaboradores (Sampietro et al., 2005) -el único estudio en el que se analizan individuos de la misma cronología y región geográfica- pertenecían a los principales haplogrupos europeos: H (52,9 %), U (17,6 %), J (11,8 %), pre-HV (5,9 %), K (5,9 %), T (5,9 %). La conclusión de este estudio fue que los antiguos íberos no eran significativamente diferentes de la población actual europeas.

Las diferencias entre nuestros resultados y los obtenidos por Sampietro y colaboradores (2005) pueden deberse bien a la existencia de una gran heterogeneidad genética entre los grupos íberos de Cataluña y los de la Comunidad Valenciana, o bien a una diferente procedencia étnica de los individuos íberos de los dos estudios. Concretamente, la elevada frecuencia del HV0 entre las muestras íberas analizadas (75 %) podría representar un carácter distintivo de nuestra población. Al no detectarse claras continuidades genéticas con las muestras del periodo precedente, el Calcolítico, podría pensarse que tuvo lugar una ruptura entre ambos periodos. Como ya se ha mencionado antes, la posibilidad de que se trate de un grupo poblacional con diferente origen genético está apoyada por el registro arqueológico, debido al tratamiento funerario diferencial (ver **Apartado D.4.1**). En el estudio de Sampietro y

colaboradores (2005) se analizaron 17 individuos adultos no incinerados procedentes de tres yacimientos de la provincia de Gerona, mientras que en la presente tesis doctoral los restos analizados procedían de la provincia de Castellón. Es posible que las diferencias en la composición genética entre ambos estudios se deban al reducido número muestral (4 individuos) de nuestra muestra o al hecho de que dependiendo de la provincia específica los supuestos enemigos ejecutados fueran de diferente origen y que hubiera conflictos entre poblaciones endogámicas, como apuntan las grandes fortificaciones de los poblados durante este periodo (ver **Apartado I.4.1.1.4**).

Desde la genética de poblaciones actuales se ha sugerido que el haplogrupo V se originó en la franja cantábrica durante el último período glacial y que desde allí se expandió hacia el norte después de la retirada de los hielos (Torroni et al., 1998, 2001), hace aproximadamente 12.000 años. A partir de entonces se habría producido una recolonización de las poblaciones aisladas en estos refugios paleolíticos hacia el noreste (Torroni et al., 1998, 2001). Sin embargo, el análisis de 121 muestras de 4 yacimientos de época Calcolítica-Bronce del País Vasco demostró la ausencia total de este haplogrupo en poblaciones vascas antiguas (Izagirre y de la Rúa, 1999). Por otra parte, un subtipaje detallado de los haplotipos y haplogrupos presentes en el grupo Saami puso de manifiesto que los linajes de mtDNA del haplogrupo V característicos de esta población estaban mucho más distribuidos en el este que en el oeste de Europa (Bermisheva et al., 2002; Tambets et al., 2004). Este resultado sugiere que el haplogrupo V habría alcanzado las áreas de Finlandia y Escandinavia a través de Europa central y del este (Tambets et al., 2004). Teniendo en cuenta todos estos datos, la presencia de este haplogrupo en el noreste de la Península Ibérica en una muestra calcolítica podría explicarse como una distribución más meridional de los refugios paleolíticos.

Otro de los haplogrupos tradicionalmente asociados al sustrato Paleolítico es el U5, cuyo tiempo de coalescencia (TMRCA) se estimó en 45.000 – 55.000 años (Richards et al., 2000). Este haplogrupo es uno de los haplogrupos más ampliamente distribuidos en Eurasia, con picos de frecuencia en la población vasca (en Guipúzcoa alcanza el 15 %) (García et al., 2011) y en Escandinavia (el U5b está representado en alrededor

del 41 % en la población Saami) (Tambets et al., 2004). También está presente en bajas frecuencias en el Norte de África (Tambets et al., 2004) y en Próximo Oriente (con frecuencias de alrededor del 2 %) (Richards et al., 2000; Tambets et al., 2004). Debido a su distribución y diferenciación actual se ha propuesto que o bien se originó en Europa o al menos fue allí dónde se diversificó su filogenia, de manera que su presencia en los otros continentes se debería a migraciones desde Europa (Richards et al., 2000). Por ejemplo, se ha sugerido que su presencia en Próximo Oriente, donde está asociada esencialmente a grupos kurdos, armenios y azerís, podría apuntar al origen europeo de los mismos (Richards et al., 2000). También el haplogrupo U5, y más concretamente el U5b, ha sido propuesto junto con los haplogrupos V, H1, H3, como un marcador de la expansión post-glacial desde los refugios paleolíticos hacia el norte de Europa, siguiendo la retirada de los hielos al final del LGM. Estas inferencias se basan en sus clinas de frecuencia, más elevadas en los supuestos refugios como la franja cantábrica (Torroni et al., 1998, 2001, 2006; Loogväli et al., 2004; Pereira et al., 2005; Alvarez-Iglesias et al., 2009).

La elevada frecuencia del haplogrupo U5 en muestras mesolíticas centroeuropeas (63 %) (Bramanti et al., 2009) y del norte de Europa (31 %) (Malmström et al., 2009) en principio apoyaría esta hipótesis. En el conjunto de muestras estudiadas este haplogrupo está presente en un individuo de época neolítica cardial (CSA24, U5*) y en uno calcolítico (COST1, U5b3). Así el individuo CSA24 podría plantear cierta continuidad genética entre el sustrato mesolítico y las primeras poblaciones neolíticas de la región. Este dato se ha visto confirmado por el reciente hallazgo de este haplogrupo en muestras mesolíticas de yacimientos del levante en estudios realizados por nuestro equipo de investigación (Gamba et al., 2011).

En lo que concierne específicamente al sub-haplogrupo U5b3, éste está actualmente poco representado en Europa, con una frecuencia por debajo del 1 % en todas las poblaciones, exceptuando Cerdeña, donde alcanza una frecuencia del 3,8 % (Pala et al., 2009). Se ha propuesto recientemente que este sub-haplogrupo se originó y se expandió a partir de un refugio paleolítico localizado en la Península Itálica siguiendo una ruta costera, puesto que los Alpes representaban una barrera geográfica (Pala et

al., 2009). Debido a la elevada endogamia sarda, consecuencia de su situación geográfica aislada, éste haplogrupo habría alcanzado una frecuencia mayor en esta región. La presencia de este mismo haplogrupo en un individuo calcolítico podría apuntar a la introducción de este haplogrupo desde las costas italianas o francesas, durante la ruta expansiva del Neolítico o en épocas posteriores.

En su conjunto, los datos analizados apuntan a cierta continuidad genética entre los diferentes períodos estudiados del noreste de la Península Ibérica, esencialmente debida a la presencia del haplogrupo K en todos ellos. Entre los resultados más interesantes merece la pena destacar las afinidades de la población neolítica con las poblaciones próximo orientales actuales, debido a la presencia de los haplogrupos N* y X1. Los resultados obtenidos también sugieren que probablemente estos haplogrupos se habrían perdido por deriva genética después del Calcolítico, período en el que su frecuencia disminuye. Por otra parte, la población íbera analizada destaca también por su composición genética diferencial, debida esencialmente a la elevada frecuencia del haplogrupo HVO. En principio esto parece sugerir que aquellos individuos no incinerados según la práctica común de la época podrían pertenecer a una población foránea, quizás enemiga, aunque el reducido número muestral analizado obliga a ser cautos con estas deducciones.

D.4.2 UTILIZACIÓN DE MÉTODOS FILOGENÉTICOS, ESTADÍSTICOS Y COMPUTACIONALES PARA LA INTERPRETACIÓN DE LOS RESULTADOS, QUE INCLUYEN LA ESTIMACIÓN DE LA CONTINUIDAD/DISCONTINUIDAD GENÉTICA ENTRE DIFERENTES PERÍODOS Y LA EVALUACIÓN DEL IMPACTO DEL NEOLÍTICO EN LA REGIÓN ESTUDIADA (ARTÍCULOS 3 Y 4)

Después de analizar la composición genética de las poblaciones estudiadas a nivel de haplotipos y haplogrupos, se consideraron las hipótesis de continuidad y discontinuidad genética entre los diferentes períodos mediante el empleo de análisis estadísticos y computacionales. Así fue posible formular las hipótesis más probables para los fenómenos demográficos de la región. Se implementaron simulaciones

bayesianas coalescentes empleando diferentes modelos de continuidad genética, para poder identificar el escenario que más se ajustaba a los resultados obtenidos.

Una de las soluciones aportadas por la estadística y la computación a los problemas derivados de la escasa disponibilidad de muestras antiguas es el empleo de simulaciones, que permiten reconstruir posibles escenarios demográficos y calcular estadísticos con los que comparar los datos reales. Además, permiten insertar los datos observados dentro de estos modelos, que emplean tamaños poblacionales cercanos a las estimaciones arqueológicas para cada uno de los períodos estudiados. Actualmente, se trata de la técnica disponible más adecuada para la correcta interpretación de los resultados.

Mediante la comparación de diferentes modelos demográficos (ver **Figura 2 del ARTÍCULO 3**), fue posible destacar cómo la ausencia de estructura en el escenario panmíctico (TPM, *Total Panmixia Model* o Modelo de Panmixia Total) era menos probable que los otros dos modelos estructurados que se consideraron (SM, *Split Model* o Modelo con separación y SDGM, *Split with Differential Growth Model* o Modelo con separación y crecimiento diferencial).

Este resultado no es sorprendente, puesto que una población panmíctica en la región estudiada desde inicios del Paleolítico hasta la actualidad no es verosímil, dado que existen movimientos poblacionales documentados en períodos prehistóricos e históricos. Sin embargo, este modelo ha sido empleado en diferentes publicaciones para evaluar el impacto del Neolítico en Europa Central (Haak et al., 2005; Bramanti et al., 2009). Por ejemplo, en el trabajo de Bramanti y colaboradores (2009), los resultados apuntaban a una ruptura genética entre los pobladores mesolíticos y neolíticos. El empleo de otros modelos de continuidad genética habrían podido explicar los datos para algunos de los valores testados, excluyendo una completa ruptura genética (Rasteiro y Chikhi, enviado a PNAS). Esto evidencia la importancia del método empleado en la interpretación de los resultados.

A raíz de eso, en la presente tesis doctoral se trabajó con tres modelos diferentes para acercarse a aquél que mejor explicaba los resultados. El modelo TPM, previamente

empleado en los trabajos citados, fue igualmente incluido para poder comparar los resultados con escenarios más verosímiles (SM y SDGM).

Para cada uno de los tres modelos, se llevaron a cabo comparaciones por pares entre las tres poblaciones antiguas estudiadas (Neolítico Antiguo, Edad de los Metales e Íberos) y los datos procedentes de la literatura de poblaciones antiguas de Neolítico Medio (Sampietro et al., 2007) y modernas (García et al., 2011) de la misma región geográfica, con el fin de evaluar los principales procesos demográficos que afectaron a la región estudiada a lo largo de la Prehistoria. La razón por la que no se analizaron conjuntamente la población del Neolítico Medio (Sampietro et al., 2007) y la del Neolítico Antiguo (de la presente tesis doctoral) se debe a que la primera incluye datos de un yacimiento en el que no se han descrito niveles cardiales subyacentes, lo que podría sesgar los resultados en el caso de no estar implicada en la llegada de los primeros agricultores en la Península Ibérica.

Obsérvese que las conclusiones del análisis descriptivo de los datos (ver **Tabla D 7**) se ven reflejadas y sostenidas por el análisis computacional. Es decir, que las distancias genéticas observadas (F_{ST}) entre las diferentes poblaciones son directamente proporcionales a las simuladas.

Tabla D 7 Distancias genéticas (F_{ST}) entre las poblaciones analizadas (**Tabla A2** del **Anexo I** del **ARTÍCULO 4**).

Edad de la Población	Moderna	Íbera	Edad de los Metales	Neolítica Media	Neolítica Antigua
Moderna					
Íbera	0,18525				
Edad de los Metales	0,02232	0,10307			
Neolítica Media	0,23213	0,18178	0,01261		
Neolítica Antigua	0,12978	0,22252	-0,02577	0,10417	

El análisis estadístico y computacional se basó en las F_{ST} por las siguientes razones:

- Entre las variables comparables y proporcionadas por los programas empleados para el cálculo de las variables observadas (Arlequin) (Excoffier y Lischer, 2010) y el cálculo de las simuladas (BayeSSC) (Excoffier et al., 2000), las F_{ST} representaban aquellas más

informativas a nivel de caracterización poblacional. En las fases preliminares del análisis estadístico se testaron otras (como la heterozigosidad) que fueron descartadas por falta de diferencias significativas.

- Se emplearon estos estadísticos también para que los resultados de la presente tesis doctoral fueran comparables con aquellos publicados por otros autores, que estudiaron problemáticas demográficas similares comparando las distancias genéticas (F_{ST}) entre los conjuntos poblacionales antiguos y modernos estudiados (Haak et al., 2005; Bramanti et al., 2009; Malmström et al., 2009).

D.4.2.1 COMPARACIÓN ENTRE DISTANCIAS GENÉTICAS OBSERVADAS Y SIMULADAS

La comparación entre las distancias genéticas observadas y simuladas permite identificar los modelos y los rangos poblacionales que más se adecúan a los datos observados. Lo que se quiere destacar es para qué valores de tamaño poblacional efectivo (N_e) neolítico y paleolítico las distancias genéticas simuladas son mayores de las observadas y con qué probabilidad. Cuando se da este fenómeno (al menos en el 5 % de los casos) en el que los valores simulados consiguen ser tan altos como los observados, se puede interpretar como buen ajuste del modelo a los datos. En el caso contrario (valores simulados menores que los observados en más del 95 % de los casos), se interpreta que los modelos y rangos poblacionales empleados no permiten explicar los resultados. La obtención de F_{ST} simulados menores que los observados en todos los rangos testados ha sido interpretada por otros autores como presencia de discontinuidad genética (Bramanti et al., 2009; Malmström et al., 2009) en aquellos casos en que el modelo testado no era capaz de explicar elevadas F_{ST} entre las poblaciones comparadas.

En los gráficos de comparación entre F_{ST} (**Figura D 3 y D 4**) se pueden distinguir fácilmente estas probabilidades gracias al código de colores empleado para representar la probabilidad de obtener F_{ST} simuladas mayores que las observadas (el color blanco se refiere a probabilidades superiores al 5 %).

En los modelos empleados se exploran diferentes tamaños muestrales paleolíticos y neolíticos. Los tamaños poblacionales efectivos durante el Paleolítico Superior y el Neolítico fueron estimados en base a los datos demográficos disponibles para estos períodos teniendo en cuenta el área geográfica objeto de estudio (noreste de la Península Ibérica). La densidad poblacional efectiva durante el Paleolítico se ha estimado en alrededor de 0,064 individuos por Km², mientras que la neolítica se supone que era veinte veces mayor (Steele et al., 1998; Alroy, 2001). Así para la región estudiada se calculó un tamaño efectivo de 700 para el Paleolítico Superior y 15.000 para el Neolítico. Pero para dar mayor flexibilidad a los modelos empleados se emplearon amplios rangos, 10 - 5.000 para el Paleolítico Superior y 200 – 100.000 para el Neolítico. Sin embargo, después de explorar estos rangos mediante un método de aproximación bayesiana (ABC), fue posible identificar un buen ajuste del rango de N_e para el Paleolítico Superior, mientras que para el Neolítico los datos apuntaban a una mayor probabilidad de la primera parte de los rangos explorados. Así se simularon F_{ST} para diferentes combinaciones y distribuciones de rangos poblacionales. Se analizaron 10 valores para el rango del Paleolítico Superior y 30 valores para el Neolítico. En este segundo caso, se distribuyeron 20 valores en la primera parte del rango (200-20.000), debido a su mayor probabilidad y 10 valores en la parte restante (20.000-100.000) (ver **Material y Métodos del ARTICULO 3**).

En la **Tabla D 8** se detalla un resumen de los resultados obtenidos en todas las comparaciones por pares, para facilitar la comprensión de la discusión de estos resultados.

Tabla D 8 Resumen de los resultados obtenidos en las comparaciones entre F_{ST} observadas y simuladas, elaborado a partir de los gráficos presentados en el presente apartado. Se especifican por cada pareja de poblaciones comparadas para qué valores es posible detectar F_{ST} simuladas mayores que las observadas en al menos el 5 % de los casos.

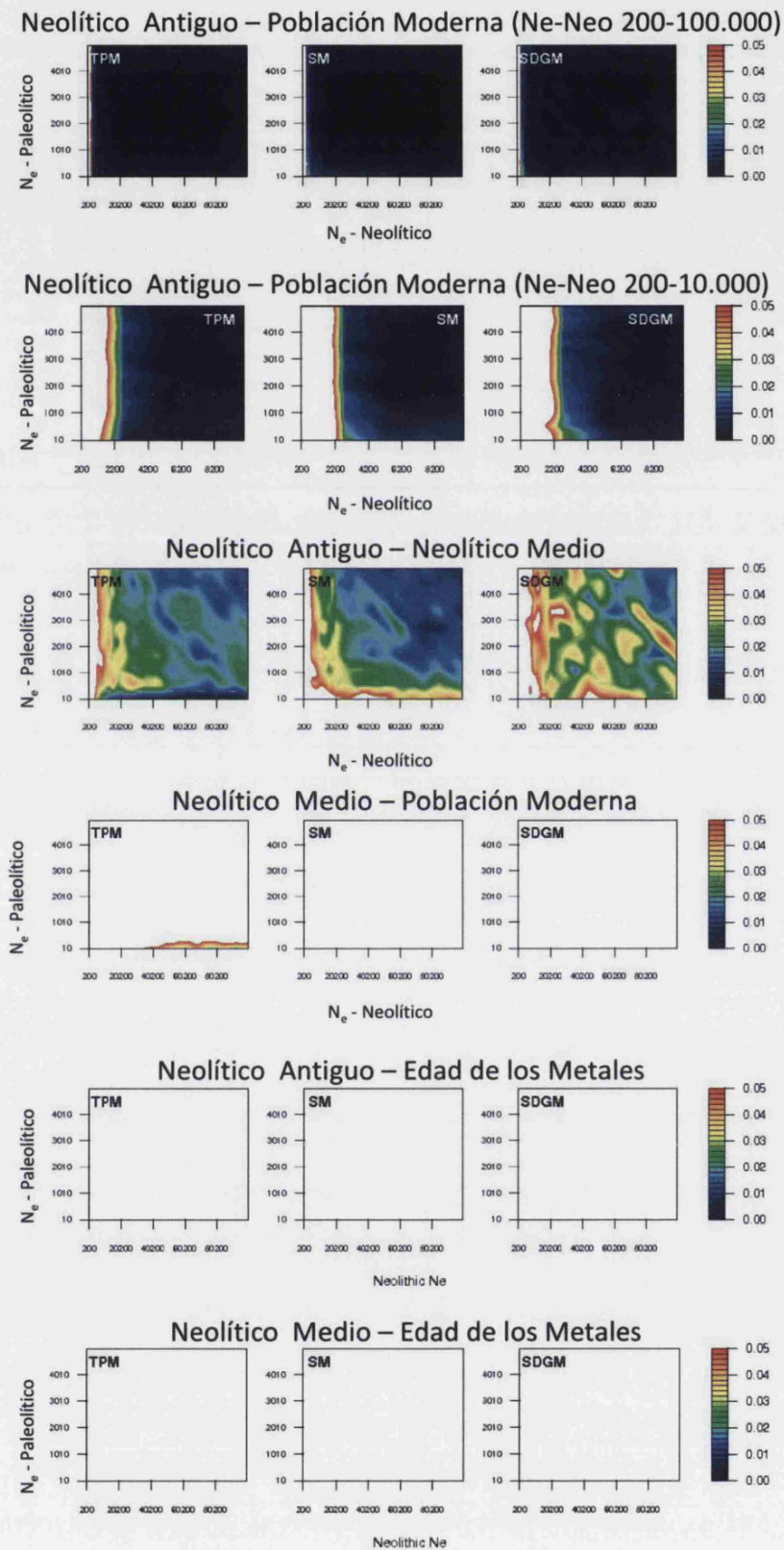
	Neolítico Antiguo	Neolítico Medio	Edad Metales	Íberos	Población actual
Neolítico Antiguo	-	-	-	-	-
Neolítico Medio	Solo para N_e neolíticos reducidos	-	-	-	-
Edad Metales	Todos los rangos	Todos los rangos	-	-	-
Íberos	Solo para N_e neolíticos muy reducidos	Solo para N_e neolíticos reducidos	Todos los rangos	-	-
Población actual	Solo para N_e neolíticos reducidos	Todos los rangos	Todos los rangos	Solo para N_e neolíticos muy reducidos	-

Este estudio comparativo ha revelado diferencias sustanciales entre la población neolítica antigua (que incluye el período cardial y post-cardial) y los demás períodos.

Respecto a su relación con la población moderna (García et al., 2011), los datos observados se pueden explicar solamente para un rango muy reducido de datos explorados en las simulaciones, compatible con un tamaño efectivo limitado de los primeros pobladores neolíticos (entre 200 y 1.000/2.000) en los tres modelos explorados (ver **Figura D 3**). Dentro de ellos, los SM y SDGM permiten explicar los resultados con tamaños poblacionales neolíticos de hasta 2.000, mientras que el TPM, solo hasta 1.000. Este resultado apoya el mejor ajuste de los modelos estructurados a los datos observados. Dado que el modelo SM refleja una aportación genética reducida de los pobladores neolíticos mientras que el SDGM refleja una aportación mucho mayor, la imposibilidad de identificar uno de estos dos modelos como más probable se debe al ajuste parecido de los dos modelos a los datos observados. Estos resultados apuntan a cierta contribución genética de los primeros agricultores, compatible con la llegada de pequeños grupos al principio del Neolítico. De ahí la interpretación demográfica de una colonización pionera de la región (ver **Discusión del ARTÍCULO 3**), que apoyaría modelos de difusión intermedios acerca de la llegada del Neolítico en la Península Ibérica (Zilhão, 2000, 2001).

También se investigó la razón por la que era difícil explicar los resultados para los demás rangos poblacionales explorados, así como el elevado F_{ST} observado entre la población neolítica antigua y la moderna. A nivel cualitativo, la población neolítica está caracterizada por una elevada frecuencia del haplogrupo N* (31 %, ver **Tabla D 5**) que, como se comentó en el **Apartado D.4.1.2**, está ausente en la población moderna (ver **Tabla S2 del ARTÍCULO 5**). Esto implica valores de F_{ST} muy elevados y difícilmente explicables bajo los tres modelos testados.

Figura D 3 Probabilidad de obtener F_{ST} simulados mayores que los F_{ST} observados entre parejas de poblaciones en los tres modelos explorados (TPM, SM, SDGM). N_e : tamaño efectivo poblacional. N_e -Neo: rango de N_e . El color blanco se refiere a valores superiores a 0.05 (5 %). Gráficos incluidos en los ARTÍCULOS 3 y 4.



Sin embargo, cuando se relaciona la población del Neolítico Medio con la población actual de la misma región geográfica, los datos obtenidos pueden explicarse para todas las combinaciones de tamaño muestral, apuntando a una homogeneidad genética entre las dos poblaciones (ver **Figura D 3**) y confirmando las conclusiones del trabajo de Sampietro y colaboradores (2007). Este resultado está reflejado incluso en el análisis cualitativo de la composición de haplotipos y haplogrupos de la población neolítica media, en la que aparecen representados los principales tipos mitocondriales actuales (ver **Tabla S2 del ARTÍCULO 3**).

En cuanto a la comparación de la población neolítica antigua con el siguiente período prehistórico (Neolítico Medio) (Sampietro et al., 2007), también se detecta cierta diferenciación genética, aunque en menor grado (ver **Figura D 3**). También es posible destacar mayores diferencias entre los tres modelos testados: los modelos estructurados, y especialmente el SDGM, pueden explicar las F_{ST} observadas para un mayor rango de combinaciones de tamaños poblacionales, apuntando nuevamente al mejor ajuste de estos modelos a los datos estudiados. Al tratarse de dos poblaciones antiguas para las que se dispone de pocos datos observados (13 para el Neolítico Antiguo y 11 para el Neolítico Medio), los modelos empleados permiten explicar los resultados por un mayor rango respecto a la comparación entre la población neolítica antigua y la moderna. Esto se debe a que fenómenos como la deriva genética pueden explicar más fácilmente cambios en la composición genética. La diferenciación genética entre Neolítico Medio y Antiguo es debida, a nivel cualitativo, a la composición haplotípica diferencial de las dos poblaciones, como ocurría con la población moderna.

En la población del Neolítico Medio están representados los principales haplogrupos europeos (ver **Tabla D 5** y **Tabla S2 del ARTÍCULO 5**), mientras que en la neolítica antigua se detectan haplogrupos no representados en estas poblaciones (X1 y N*, ver **Tabla D 5**). Este resultado se podría explicar en el caso de que la deriva genética hubiera jugado un papel importante, que acarrearla la pérdida de algunas variantes mitocondriales y la modificación de las frecuencias haplotípicas. Esto explicaría la pérdida del haplogrupo N* en la población moderna y en la del Neolítico Medio. El

hecho de que se encuentre un haplogrupo N* en una muestra calcolítica apunta a su persistencia al menos hasta ese periodo, aunque en menor frecuencia. Por otro lado, su ausencia en el conjunto de muestras del Neolítico Medio (Sampietro et al., 2007) podría ser debido a que no se trata de un conjunto representativo, puesto que incluye 11 individuos procedentes de un solo yacimiento, que además no presenta niveles neolíticos cardiales.

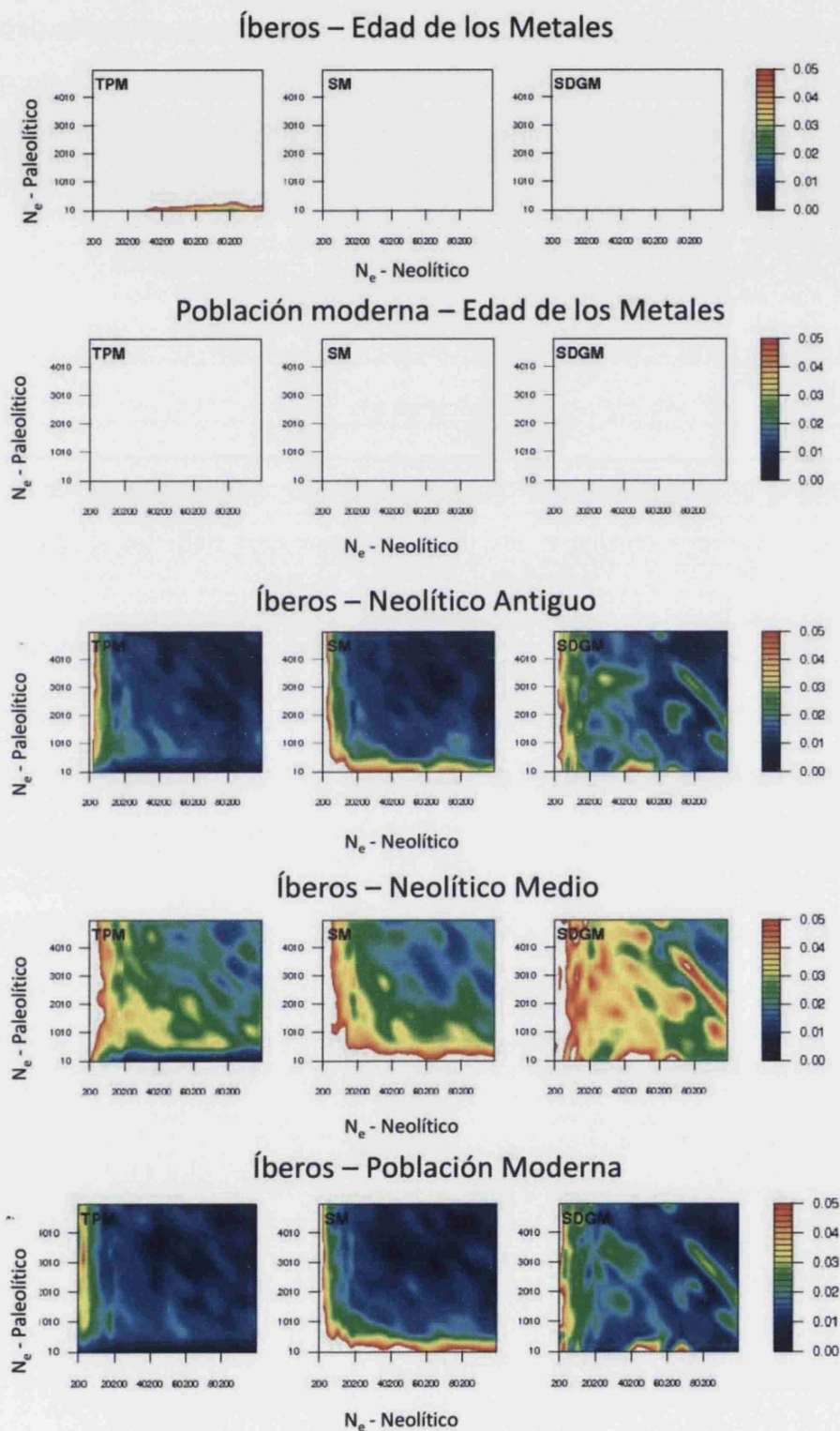
Se comparó también la población neolítica antigua con las poblaciones calcolítica y de la Edad del Bronce (conjuntamente definidas como Edad de los Metales). En este caso puede verse que es posible explicar los resultados a nivel de todos los rangos de población estudiados en todos los modelos ensayados (ver **Figura D 3**). A nivel cualitativo, ambos grupos poblacionales comparten diferentes haplotipos y haplogrupos. Por ejemplo, en ambos casos el haplogrupo K está representado en más de un individuo y la población calcolítica también presenta un individuo caracterizado por el haplogrupo N*. Estos resultados apuntan a la existencia de continuidad genética entre el Neolítico Antiguo y los sucesivos periodos prehistóricos (Calcolítico y Edad del Bronce).

Por otra parte, se observa una situación parecida cuando se compara la Edad de los Metales con todos los restantes periodos estudiados (ver **Figura D 3 y D 4**), sugiriendo homogeneidad genética entre todos estos periodos. Estos resultados parecen discordantes con los anteriores, ya que entre el Neolítico Antiguo y Medio así como entre el Neolítico Antiguo y la población moderna se detecta cierta diferenciación genética. Sin embargo, ambas poblaciones resultan genéticamente homogéneas con la población de la Edad de los Metales (ver **Figura D 3**). Estos resultados se pueden explicar debido a los bajos valores de F_{ST} obtenidos entre la población de la Edad de los Metales y estas poblaciones (Neolítico Antiguo, Medio y población actual, ver **Tabla D 7**). Así la probabilidad de obtener valores simulados mayores respecto a los observados resulta altamente probable independientemente del modelo. Ahora bien, lo que resulta más complicado de explicar es la razón por la que estos valores de F_{ST} son tan bajos. Por un lado la población de la Edad de los Metales presenta un número reducido de individuos (8) pero no mucho más reducido que las poblaciones neolíticas

media y antigua (11 y 13 individuos respectivamente). La razón más probable puede residir en la manera en la que se calculan los F_{ST} , que esencialmente se basan en las diferencias mutacionales entre los haplotipos comparados. Así su valor disminuye en función de las mutaciones compartidas entre los haplotipos de las poblaciones comparadas. La población de la edad de los metales incluye haplotipos compartidos o parcialmente compartidos con las otras tres poblaciones. Sin embargo, dependiendo de la comparación, comparte unas u otras mutaciones (ver **Tabla D 5** y **D 6**). Por ejemplo, respecto al Neolítico Antiguo presenta dos haplotipos coincidentes (16224C 16311C y 16223T 16362C) y otro parcialmente coincidente (16136C, 16192T 16270T (16304C)). Respecto a la población del Neolítico Medio comparte dos haplotipos (CRS y 16069C 16126C). Por lo que respecta la población moderna, comparte con ella los haplotipos que actualmente están mayormente representados (CRS: 40,4 %, 16224C 16311C: 6,6 %, 16298C: 5,4 % y 16069C 16126C: 5,4 %).

En lo que respecta al conjunto de muestras íberas, esta población presenta una elevada diferenciación genética con todas las demás poblaciones, exceptuando la Edad de los Metales (ver **Figura D 4**). Como ya se comentó en el anterior apartado, 3 de las 4 muestras estudiadas se corresponden con el haplogrupo HV0. Así existen dos posibilidades que podrían explicar los resultados obtenidos. En primer lugar, habría que considerar si la población íbera estudiada es representativa de las poblaciones íberas del momento. Por una parte, hay que tener en cuenta que únicamente se estudiaron 4 muestras, lo que podría haber sesgado el resultado de alguna manera. Por otra, y como ya se ha discutido en el apartado anterior, los individuos analizados presentaban un ritual funerario diferente al de las poblaciones íberas de la Península. Como ya se ha apuntado, esto podría significar que se tratara de una población alóctona enemiga.

Figura D 4 Probabilidad de obtener F_{ST} simulados mayores que los F_{ST} observados entre parejas de poblaciones, en correspondencia de los tres modelos explorados (TPM, SM, SDGM). N_e : tamaño efectivo poblacional. Ne-Neo: rango de N_e . El color blanco se refiere a valores superiores a 0.05 (5 %). Gráficos incluidos en el **ARTÍCULO 3 y 4**.



Por otra parte, en segundo lugar, podría ser que la población estudiada sí fuera representativa de las poblaciones ibéricas, por lo que las elevadas distancias genéticas observadas y simuladas se deberían a una composición genética diferencial de esta población. Sin embargo esta segunda hipótesis resulta menos probable debido a la diferente composición genética que presenta respecto a otro conjunto de muestras ibéricas estudiadas por otros autores (Sampietro et al., 2005), en la que no está representado ninguno de los haplotipos ni de los haplogrupos encontrados en el conjunto íbero de la presente tesis doctoral.

D.4.2.2 SELECCIÓN DEL MODELO MÁS ADECUADO

Si se observan todos los gráficos en su conjunto, es importante destacar que los modelos estructurados consiguen explicar los resultados para un rango mayor de poblaciones efectivas a nivel de todas las comparaciones efectuadas. A continuación se describen los análisis estadísticos que apoyan el mejor ajuste de los modelos SM y SDGM al conjunto de datos estudiados.

Mediante los análisis de ABC (*Approximate Bayesian Computation*), fue posible demostrar que los modelos que asumen una estructura previa de las poblaciones Paleolítica y Neolítica (SM y SDGM) son más probables que el modelo sin estructura (TPM) (ver las probabilidades correspondientes en la **Tabla 2** del **ARTÍCULO 3**). Además, si se lleva a cabo una comparación por pares entre los modelos, se obtienen probabilidades a posteriori más elevadas (>72 %) para los escenarios estructurados.

Con los resultados obtenidos en la presente tesis doctoral no fue posible discernir entre estos dos modelos demográficos por diferentes razones. En primer lugar, los datos genéticos empleados pertenecen, en su totalidad, a épocas posteriores a la estructura que se ha diseñado en los modelos, es decir que son más recientes de hace 7.500 años. El empleo de datos que precedan a esta fecha permitiría afinar la selección del modelo más adecuado. Teniendo en cuenta que la rama izquierda de los modelos estructurados representa la población local y la derecha la población fuente (ver **Figura 2** del **ARTÍCULO 3**), sería necesario añadir datos genéticos del Mesolítico del noreste

de España e información de la población neolítica de Próximo Oriente. En ambos casos se trataría de datos que estarían incluidos exactamente en el período empleado para estructurar los modelos, es decir entre hace 10.000 y 7.500 años.

Al no disponer de estos datos, se llevó a cabo una prueba en la que la edad de la estructura se redujo a 7.395 años, de manera que las siete muestras de Can Sadurní quedaran justo antes de la misma (en la rama izquierda, como si representaran el sustrato mesolítico) y las de Chaves y Sant Pau del Camp justo después. Los resultados de selección de modelos en este caso mejoraron sensiblemente, identificando como modelo más adecuado el SDGM asociado a una probabilidad superior al 90 %.

Para validar los resultados y las conclusiones obtenidas, se estimó el poder discriminatorio del método empleado para la selección del modelo más probable. Se seleccionaron 500 valores al azar dentro del conjunto de datos simulados (1 millón por modelo). Estos valores se consideraron como si fueran observados y se calculó la probabilidad de asignación a cada uno de los modelos, evaluando cuántas veces se correspondía esta probabilidad con el modelo correcto (ver **Tabla D 9**).

Tabla D 9 Probabilidad de identificar el modelo de procedencia seleccionando 500 valores al azar dentro del conjunto de datos simulados (1 millón) por cada uno de los modelos: TPM (primera línea), SM (segunda línea) y SDGM (tercera línea) (**Tabla S4** del **ARTÍCULO 3**).

% de atribución a cada modelo			
	TPM	SM	SDGM
TPM	60,20	15,06	24,20
SM	45,40	26,00	28,60
SDGM	31,40	24,80	43,80

En orden, los datos procedentes del TPM se asignaron correctamente a este modelo en el 60 % de los casos (y en el restante 40 % a los otros dos modelos), aquellos procedentes del SDGM en un 44 % y en el SM solamente en un 26 %. Esto es debido a las mayores diferencias entre TPM y los otros dos modelos que los dos modelos estructurados (SM y SDGM) entre sí. En el caso de TPM y SDGM, los porcentajes más altos de asignación corresponden al modelo correcto. En el caso del SM, aunque el porcentaje más alto cae dentro del modelo TPM, este porcentaje no se diferencia significativamente de los otros dos (26,00 % y 28,60 %, ver **Tabla D 9**). Este resultado

podría ser debido a que el SM representa un modelo intermedio entre TPM y SDGM, lo que conllevaría un porcentaje de error más elevado en la identificación del modelo.

En su conjunto, todos estos resultados apuntan a que los modelos estructurados (SM y SDGM) son mucho más compatibles con los datos reales respecto a modelos panmícticos, poco verosímiles incluso desde un punto de vista conceptual. Como ya se ha comentado, el modelo TPM fue incluido en este estudio para poder comparar los resultados con los de otros autores (Haak et al., 2005; Bramanti et al., 2009; Malmström et al., 2009), que lo emplearon como único modelo para testar continuidad genética. Futuras investigaciones para la identificación de modelos más adecuados y la inserción de nuevos datos podrán ayudar a identificar más claramente el modelo demográfico más adecuado para comprender las aportaciones del Neolítico a la genética actual.

DISCUSIÓN DE LOS OBJETIVOS – RELACIONES FAMILIARES

D.5 Verificación de las hipótesis de parentesco y de determinación del sexo formuladas a partir de otras disciplinas (arqueología, historiografía, genealogía, antropología) mediante el análisis genético de marcadores nucleares STR en un conjunto de muestras medievales procedentes de sepulturas múltiples (Iglesia de San Esteban de Cuéllar, Segovia).

Durante el desarrollo de la presente tesis se ha demostrado la posibilidad de obtener perfiles de STRs nucleares prácticamente completos en muestras medievales mediante el empleo del *kit MiniFiler* (Mulero et al., 2008). La obtención de resultados tan satisfactorios probablemente se deba no solo a la alta sensibilidad del *kit*, sino también y sobre todo a la muy buena preservación de las muestras. No hay que olvidar que una de las principales características de este conjunto de muestras es su elevado grado de preservación tisular. No en vano se trataba de especímenes momificados de forma natural. Este proceso de momificación natural depende del entorno de deposición y se trata de un fenómeno que se ve favorecido por diferentes factores, como una desecación rápida de los tejidos incluso en ambientes cálidos, un ambiente alcalino o la ausencia de microorganismos (Lynnerup, 2007). En el caso de las muestras estudiadas, la momificación fue favorecida por un ambiente seco, proporcionado por los sepulcros de piedra en los que se inhumaron los cuerpos. Además, en el interior de algunos de los sepulcros se encontró una gran cantidad de cal, una sustancia que absorbe la humedad ambiental, favoreciendo aún más el fenómeno de momificación natural de los cuerpos.

D.5.1 INTERPRETACIÓN DE LOS PERFILES STRS (ARTÍCULO 5)

El éxito de este estudio fue condicionado, además de por las buenas condiciones de preservación de las muestras, por la correcta interpretación de los perfiles genéticos. A

la hora de llevar a cabo estudios de perfiles de STRs con muestras degradadas o LTDNA, en inglés *Low Template DNA* (bajo número de ADN molde), se suele presentar un conjunto de problemáticas específicas que las distinguen del análisis e interpretación de muestras de ADN fresco (Balding y Buckleton, 2009; Gill et al., 2009). Por esta razón, la correcta lectura e interpretación de los perfiles representa el punto de partida para estudios de este tipo. En primer lugar, los picos correspondientes a la amplificación de los diferentes sistemas deben tener una altura (medida en RFU, *Relative Fluorescence Units* o Unidades Relativas de Fluorescencia) lo suficientemente elevada como para no confundirla con el ruido de fondo. Siguiendo las recomendaciones vigentes, no se consideraron como válidos los picos por debajo de las 50 RFU (Gjertson et al., 2007).

Además, resulta importante la identificación de posibles artefactos ligados a la baja concentración de ADN. En el presente estudio se identificaron el desequilibrio alélico en heterocigotos (*allelic imbalance*), la pérdida aleatoria de información genética de uno de los alelos de un locus (*allelic dropout*) y la pérdida de información genética de un locus completo (*locus dropout*). Sin embargo existen otros artefactos no detectados en el presente trabajo que podrían afectar a la interpretación de los resultados de muestras LTDNA, como la presencia de amplicones de altura reducida en correspondencia de una unidad de repetición inferior a la verdadera (llamados *stutter*) y la detección de contaminaciones puntuales y no replicables, conocido como *drop-in* (Gill et al., 2000; Whitaker et al., 2001; Kloosterman y Kersbergen, 2003; Budowle et al., 2009).

En las muestras medievales de la Iglesia de San Esteban, se observa el fenómeno del *allelic imbalance* repetidamente. Este artefacto consiste en el desequilibrio en la altura de los picos del electroferograma en sistemas heterocigotos. En muestras frescas y correctamente amplificadas, la altura del pico de menor tamaño suele ser aproximadamente el 90 % de la del pico de mayor tamaño (Gill et al., 1997). Aunque existen diferentes estándares en los diferentes laboratorios forenses, en general se suele requerir una relación de al menos el 60 % (Gill et al., 1997). En el caso de las muestras de la Iglesia de San Esteban, la altura de los picos de menor tamaño llegó a

representar incluso el 30 % de la altura de los picos de mayor tamaño. Para poder confirmar que el pico de menor tamaño no fuera un artefacto, se analizaron los perfiles obtenidos en las diferentes amplificaciones y extracciones, hasta llegar a un consenso. Además se excluyó la posibilidad de que se tratara de *stutters* debido a que en amplificaciones independientes el pico de menor altura se presentaba alternativamente a nivel del alelo de mayor o menor tamaño.

La razón por la que se puede presentar el fenómeno del *allelic imbalance* se debe a la degradación y a la baja concentración de ADN, que pueden conllevar la amplificación preferente de un alelo frente al otro (Hummel, 2003; Butler, 2005). También puede ocurrir en muestras frescas en el caso de que uno de los dos alelos presente una mutación en la región de hibridación de los cebadores, provocando un decremento en la eficiencia de amplificación del alelo en cuestión (Butler, 2005). No obstante, los cebadores empleados en este tipo de análisis suelen hibridar en zonas con una probabilidad de mutación muy reducida (Clayton et al., 1998).

También fue posible detectar el fenómeno del *allelic dropout*, que consiste en la amplificación al azar de un único alelo en una muestra heterocigota para un determinado marcador (Butler, 2005). A priori, esto podría parecer una situación extrema del fenómeno de *allelic imbalance*, puesto que el resultado es la detección de un solo pico en un *locus* heterocigoto. Sin embargo, Findlay y colaboradores (Findlay et al., 2001) sugirieron que se trataba de dos fenómenos diferentes. Estos autores afirman que a la hora de preparar una PCR se añade solo una porción del extracto de ADN. Si se está trabajando con muestras degradadas, la concentración de ADN en el extracto podría ser tan baja que el volumen añadido a la mezcla de PCR solo contenga uno de los dos alelos.

Como en el caso anterior, gracias a la repetición de los experimentos fue posible detectar este fenómeno para poder establecer el haplotipo consenso. Es decir, que en aquellas ocasiones en las que se obtuvieron dos alelos distintos, en amplificaciones diferentes y para el mismo marcador, se asumió que se trataba de *allelic dropout* (ver **Tabla D 10**).

Tabla D 10 Perfiles de las muestras amplificadas con el kit *MiniFiler*. Los alelos con un asterisco (en las celdas grises) se refieren al fenómeno del *allelic dropout* y la presencia de un guión, a la ausencia de resultados (*locus drop-out*).

	Individuo	Muestra	PCR	D13S317	D7S820	AMEL	D2S1338	D21S11	D16S539	D18S51	CSF1PO	FGA
Tumba A	1SC	1SC1	1	12/13	-	-	18/19	-	-	13/14	10/11	-
			2	-	-	-	-	-	-	11*	-	
		1SC2	1	13*	-	X*	18/19	-	-	13*	10/11	-
			2	12*	-	X/Y	18/19	-	-	13/14	10/11	-
	Perfil consenso			12/13	-	X/Y	18/19	-	-	13/14	10/11	-
	2SC	2SC1	1	11/12	-	X/Y	19*	-	-	13/14	10/11	-
			2	11/12	-	X/Y	17/19	-	-	13/14	10/11	-
		2SC2	1	11/12	-	-	-	-	-	13*	10/11	-
			2	-	-	-	-	-	-	13*	10/11	-
	Perfil consenso			11/12	-	X Y	17/19	-	-	13/14	10/11	-
	3SC	3SC1	1	10/12	8/9	X/Y	19/24	28/32.2	12/13	11/18	10/11	23/24
			2	10/12	8/9	X/Y	19/24	28/32.2	12/13	11/18	10/11	23/24
		3SC2	1	10/12	8/9	X/Y	19/24	28/32.2	12/13	11/18	10/11	23/24
			2	10/12	8/9	X/Y	19/24	28/32.2	12/13	11/18	10/11	23/24
	Consensus profile			10/12	8/9	X/Y	19/24	28/32.2	12/13	11/18	10/11	23/24
Tumba B	4SC	4SC1	1	11	8	X	24*	-	-	-	11	22
			2	11	-	X	-	28/30	-	15/16	11	22
		4SC2	1	11	-	X	20*	-	-	13/16	11	-
			2	11	-	X	20/24	-	-	13/16	11	22
	Perfil consenso			11/11	8/8	X/X	20/24	28/30	-	13/16	11/11	22/22
Tumba C	5SC	5SC1	1	9/11	10/12	X	17/25	28/30	13	14/17	10/12	22/23
			2	9/11	10/12	X	17/25	28/30	13	14/17	10/12	22/23
		5SC2	1	9/11	10/12	X	17/25	28/30	13	14/17	10/12	22/23
			2	9/11	10/12	X	17/25	28/30	13	14/17	10/12	22/23
	Perfil consenso			9/11	10/12	X/X	17/25	28/30	13/13	14/17	10/12	22/23
	6SC	6SC1	1	10/11	8/10	X/Y	19/24	28/32.2	9/12	17/18	11/12	24
			2	10/11	8/10	X/Y	19/24	28/32.2	9/12	17/18	11/12	24
		6SC2	1	10/11	8/10	X/Y	19/24	28/32.2	9*	17/18	11/12	24
			2	10/11	8/10	X/Y	19/24	28/32.2	9*	17/18	11/12	24
	Perfil consenso			10/11	8/10	X/Y	19/24	28/32.2	9/12	17/18	11/12	24/24
Tumba D	7SC	7SC1	1	11/13	8/12	X/Y	24/25	29/32	9	12/15	11/12	19/23
			2	11/13	8/12	X/Y	24/25	29/32	9	12/15	11/12	19/23
		7SC2	1	11/13	8/12	X/Y	24/25	29/32	9	12/15	11/12	19/23
			2	11/13	8/12	X/Y	24/25	29/32	9	12/15	11/12	19/23
		7SC3	1	11/13	-	X/Y	24/25	29/32	-	12/15	12*	19*
			2	-	-	-	-	-	-	-	12*	-
	Perfil consenso			11/13	8/12	X/Y	24/25	29/32	9/9	12/15	11/12	19/23

Por ejemplo, la muestra 1SC2 presenta para el marcador D13S317 los alelos 12 y 13 en dos amplificaciones independientes, interpretándose que, debido al *allelic dropout*, en cada PCR se ha amplificado uno de los dos alelos presentes en la muestra. Esto hace que dudemos de todos aquellos sistemas genéticos que muestren un solo alelo, ya que

cabe la posibilidad de que otro alelo distinto, presente en la muestra, no se haya amplificado. Por esta razón la repetición de los análisis mediante extracciones independientes de diferentes muestras de un mismo individuo y diferentes amplificaciones a partir del mismo extracto es de vital importancia para una correcta interpretación de los resultados obtenidos. En el caso del individuo 1SC es posible ver que entre las dos amplificaciones llevadas a cabo para cada una de las dos muestras se han podido tipar entre 1 y 5 sistemas por PCR, apuntando a la escasez de ADN en la muestra.

El tercer artefacto más frecuente en muestras LTDNA es el *locus dropout*, o pérdida de información de un *locus*. Dicho fenómeno también fue detectado en el conjunto de muestras estudiadas. A pesar de obtener perfiles genéticos parciales en algunas de las amplificaciones, fue posible reconstruir el haplotipo completo como es el caso de la muestra 7SC. En cambio, para los individuos 1SC, 2SC y 4SC el perfil consenso es a su vez parcial debido a que no fue posible amplificar algunos marcadores en ninguna de las amplificaciones realizadas (ver **Tabla D 10**). Los sistemas principalmente afectados por el *locus dropout* son aquellos que presentan una mayor longitud (FGA, D21S11 y D7S820), sin embargo este fenómeno también afectó a un sistema de reducido tamaño (D16S539), por el que se han descrito mutaciones a nivel de la región de unión del primer específico (Feller, 2009). En la **Figura S2 (Material Suplementario del ARTÍCULO 5)** se detallan las diferentes longitudes de cada uno de los fragmentos analizados.

D.5.2 INTERPRETACIÓN DE LOS RESULTADOS GENÉTICOS OBTENIDOS A LA LUZ DE HIPÓTESIS ARQUEOLÓGICAS, VALORÁNDOLOS CONJUNTAMENTE CON LA INFORMACIÓN GENEALÓGICA, HISTÓRICA, ARQUEOLÓGICA Y ANTROPOLÓGICA DISPONIBLE (ARTÍCULO 5)

En su conjunto, este trabajo representa un ejemplo de una de las aplicaciones del ADN antiguo: la determinación de relaciones familiares en un contexto histórico y arqueológico. Mediante la integración de información procedente de diferentes disciplinas, que incluyen la arqueología, la antropología, la genealogía y la genética, fue

posible llevar a cabo un estudio de relaciones de parentesco en un conjunto de sepulcros medievales.

Los siete individuos analizados proceden de un grupo de sepulcros de interés histórico, puesto que estaban localizados en posiciones privilegiadas de la Iglesia de San Esteban de Cuéllar, reservadas a miembros de familias nobles durante la Edad Media. De hecho, las inscripciones de las tumbas hacían referencia a personalidades de la familia Córdoba Hinestrosa, probablemente emparentada con Alfonso IX de León (Palomino Lázaro et al., 2009). De ahí el gran interés en el estudio de las relaciones familiares de los siete individuos hallados en los cuatro sepulcros mediante un análisis multidisciplinar. Si en un futuro fuera posible obtener la información genética de descendientes vivos (o muertos), sería posible llegar a identificar a los individuos estudiados.

Los hallazgos más significativos de este trabajo se refieren a dos aspectos:

- La determinación molecular del sexo y la comparación de los resultados con las inscripciones y los datos antropológicos.
- El análisis de parentesco entre los individuos supuestamente relacionados.

D.5.2.1 DETERMINACIÓN MOLECULAR DEL SEXO

La determinación molecular del sexo resulta de gran interés sobre todo en el caso de no disponer de un esqueleto completo o de individuos inmaduros. En ambos casos, la determinación morfométrica del sexo se ve imposibilitada por la falta de información. Sin embargo, la genética molecular puede permitir su determinación, como es el caso de los dos individuos infantiles estudiados. Además, el gen de la amelogenina, incluido en el *kit MiniFiler*, presenta un tamaño reducido, permitiendo fácilmente su amplificación en restos degradados.

Como puede verse en la **Tabla D 10** y en **Tabla 3** del **ARTÍCULO 5** fue posible determinar el sexo de todos los individuos estudiados. En primer lugar, se compararon los resultados obtenidos con las determinaciones realizadas por el equipo antropológico mediante métodos morfométricos. Los resultados genéticos fueron consistentes con los antropológicos en cuatro de los cinco individuos adultos. Sin embargo, para el esqueleto 5SC, clasificado antropológicamente como varón, se pudo amplificar tan solo el marcador correspondiente al cromosoma X. Teniendo en cuenta que para cada uno de los extractos de ADN de este individuo fue posible amplificar repetidamente todos los sistemas incluidos en el *kit MiniFiler*, se excluye la posibilidad de que la amplificación del cromosoma X represente un *allelic dropout*.

Por otra parte, la probabilidad de que un hombre presente una deleción en el fragmento de la amelogenina amplificado en su cromosoma Y y por tanto sea clasificado erróneamente como mujer es muy reducida. Esta frecuencia fue estimada en una amplia base de datos de población europea en un 0,02 % (Steinlechner et al., 2002). A favor de la identificación del esqueleto 5SC como mujer, está la inscripción de la tumba, que cita “Doña Urraca García de Tapia”, por lo que cabría esperar su presencia en el sepulcro. Así los resultados obtenidos mediante el estudio genético permitieron la corrección del sexo de este individuo y la corroboración de la hipótesis histórica basada en la inscripción de la tumba.

D.5.2.2 RELACIONES FAMILIARES

La comparación entre resultados genéticos e hipótesis arqueológicas permitió evaluar las posibles relaciones de parentesco entre los individuos estudiados. A partir de los haplotipos consenso se calcularon las LR (razones de verosimilitud, en inglés *Likelihood Ratios*) y W (probabilidad condicionada a posteriori, calculada a partir del LR, asumiendo la igualdad de las dos hipótesis – a priori = 0,5) (Gjertson et al., 2007) con un programa (FAMILIAS) específicamente diseñado para el estudio de *pedigrees* alternativos y ampliamente empleado en genética forense (Egeland et al., 2000).

Uno de los intereses arqueológicos era identificar los posibles padres de los dos individuos infantiles (1SC y 2SC). Para ambos solo fue posible amplificar y tipar 4 de los 8 sistemas autosómicos (ver **Tabla D 10**), reduciendo de manera significativa las probabilidades de filiación asociadas. En primer lugar, dado que estaban enterrados junto con el individuo masculino 3SC, se investigó la posibilidad de que se tratara de su padre. Sin embargo presentaba con ambos niños una exclusión en el sistema D18S51, con una distancia mínima de dos pasos mutacionales (entre alelo 11 y 13, ver **Tabla D 10**), que se interpreta como probabilidad de paternidad nula. También existe la posibilidad de que se tratara de hijos ilegítimos y por ello no es posible encontrar evidencias genéticas.

Se suponía que los individuos infantiles eran hermanos y que si no eran hijos del individuo 3SC, podrían ser hijos de alguno de los otros individuos enterrados en los otros tres sepulcros y enterrados o movidos al sepulcro A sucesivamente. En este caso, los perfiles de 1SC y 2SC compartían al menos un alelo en los 4 sistemas tipados (ver **Tabla D 10**), por lo que la probabilidad de hermandad alcanzó un 68 % a pesar del escaso número de marcadores analizados.

No fue posible encontrar ningún individuo adulto que fuera compatible como padre o madre de ambos niños. Sin embargo, el individuo 5SC no presentaba ninguna exclusión con uno de ellos (2SC, ver **Tabla D 10**) y se calculó una probabilidad de maternidad del 68 %. El perfil genético de esta mujer no era compatible con el del otro niño, presentando exclusiones en el 50 % de los marcadores analizados (2 sobre 4).

Una probabilidad de hermandad o maternidad de alrededor del 68 %, como la que se ha obtenido en el presente estudio (1SC / 2SC y 5SC / 2SC), no se admitiría en una prueba estándar forense, sobre todo si se tiene en cuenta que hoy en día se obtienen valores de probabilidad superiores en varios órdenes de magnitud ($LR > 100.000$ y $W > 99.999 \%$) (Buckleton et al., 2004). Sin embargo, dado que se trata de muestras arqueológicas en las que resulta difícil recuperar ADN, se hace necesario valorar la información genética como un dato más en el contexto general de la información global del estudio.

A pesar de los bajos porcentajes obtenidos, estos resultados podrían apoyar diferentes hipótesis arqueológicas: (1) 1SC y 2SC podrían tener padres diferentes y la madre de 2SC podría ser 5SC con una probabilidad del 68 %; (2) 1SC y 2SC podrían ser hermanos con una probabilidad del 68 %, pero en este caso ninguno de los padres estaría enterrado en el conjunto de tumbas analizadas. Sin embargo, los datos arqueológicos disponibles no nos permiten decantarnos por ninguna de las dos hipótesis.

Otra de las hipótesis arqueológicas más firmes, se refería a la posible relación genética paterno-filial entre los individuos 5SC y 6SC, debido a su enterramiento conjunto (sepulcro C) y a la coincidencia entre la edad de la muerte de los dos individuos. Los resultados genéticos no confirmaron esta posibilidad puesto que se encontraron 3 exclusiones sobre 8 sistemas comparados (ver **Tabla D 10**). No obstante, para los individuos 3SC y 6SC, que presentaban perfiles completos y estaban enterrados en dos tumbas diferentes (A y C respectivamente), se identificó una relación paterno-filial asociada a una elevada probabilidad (99,98 %), que los sitúa incluso por encima del límite para una prueba de paternidad establecido por la jurisprudencia española ($W=99,75$ % correspondiente a un $LR= 400$, a priori de 0.5, sentencia nº 593/1992 del Tribunal Supremo, 1992). También en este caso existía cierta sospecha arqueológica de esta relación, puesto que en el sepulcro C se podía intuir una inscripción poco clara que parecía referirse a la presencia de Gabriel López de Córdoba-Hinestrosa en su interior. Según el archivo histórico, este último fue el padre de Martín López de Córdoba-Hinestrosa (3SC), enterrado en la tumba A. El análisis genético permitió confirmar esta hipótesis arqueológica. Los valores obtenidos para el índice de paternidad entre 3SC y 6SC y la probabilidad de paternidad fueron respectivamente 4422.24 y 99.98 %, comparables con los valores que se obtienen en el campo forense.

Para la interpretación de los resultados hay que tener en cuenta que a la hora de calcular las probabilidades es necesario conocer las frecuencias alélicas de la población a la que pertenecen los individuos estudiados. Estas frecuencias se emplean en el denominador para calcular la probabilidad de coincidencia por azar. Sin embargo, el presente caso no existe información de la frecuencia y distribución de estos alelos, dado que los individuos estudiados pertenecen a poblaciones pretéritas, cuyas

distribuciones de frecuencias no tienen por qué coincidir con las actuales. Dada la imposibilidad de obtener estos datos, se emplearon aquellos correspondientes a población actual general española. Al utilizar las frecuencias de una población actual, estamos suponiendo que ha habido una continuidad genética en estas poblaciones desde el período estudiado hasta la actualidad. Este hecho no está probado, por lo tanto es muy posible que en el cálculo se cometa un error, ya sea por sobrestimación o por infraestimación de las frecuencias, error del que no podemos controlar ni su magnitud ni su posible efecto. Otra cuestión relacionada en cierta manera con lo anterior es nuestro desconocimiento del grado de endogamia de los grupos humanos a los que pertenecían estos individuos. Un alto grado de endogamia hace más probable que los individuos compartan alelos por azar y esto haría que estuviésemos sobrevalorando los índices de paternidad, maternidad o hermandad.

CONCLUSIONES

CONCLUSIONES (ESPAÑOL)

CONCLUSIONES - METODOLOGÍA Y VALIDACIÓN DE LOS RESULTADOS

Métodos experimentales

1. El análisis comparativo de diferentes métodos de extracción de ADN ha permitido seleccionar e implementar en el laboratorio el protocolo diseñado por Rohland y Hofreiter (2007), basado en el empleo de sílica, en sustitución del protocolo habitualmente empleado (basado en el empleo de Fenol/Cloroformo), puesto que está asociado a una elevada eficiencia en la obtención de resultados replicables. Este protocolo elimina eficazmente las moléculas inhibitoras, evita la exposición a sustancias tóxicas volátiles, permite solapar diferentes tandas de extracción, filtrar las soluciones y emplear material estéril en todo momento, es escalable y, en el caso de muestras bien preservadas, el GuSCN puede ser sustituido por cloruro sódico (NaCl), abaratando considerablemente el coste del protocolo.
2. Se ha demostrado la elevada eficiencia del *Multiplex PCR kit* de QIAGEN en la amplificación de aDNA, aunque su elevada sensibilidad también se asocia a una amplificación de moléculas contaminantes en el caso de muestras muy degradadas o con bajo contenido de ADN. Permite, sin embargo, amplificar muestras con alto contenido de inhibidores, por lo que el costo-beneficio de todos estos factores es claramente positivo.
3. Ha sido posible obtener información genética de marcadores nucleares a partir de muestras antiguas mediante la utilización de herramientas diseñadas para el campo forense, como el *AmpF&STR® MiniFiler™ PCR Amplification Kit* de Applied Biosystems, aunque el éxito en la amplificación está sujeto a la buena preservación de las muestras. Así mismo, el software

empleado ha sido de utilidad para aceptar y/o rechazar hipótesis arqueológicas de parentesco.

4. Con los protocolos utilizados ha sido posible obtener información genética reproducible a partir de muestras de más de 7.000 años de antigüedad de entornos templados.

Prospección

5. La procedencia de la muestra es la variable que más influye en la recuperación de ADN a partir de muestras arqueológicas. El entorno de deposición parece jugar un papel fundamental en la preservación molecular del ADN, más importante que la antigüedad de la muestra.
6. La cuantificación espectrofotométrica del ADN mediante la absorbancia de UV no refleja el contenido de ADN amplificable ni el contenido de moléculas inhibitoras, revelándose como una medida no apta para estudios prospectivos en el campo del ADN antiguo.

Validación de los resultados

7. La contaminación detectada mediante amplificación reiterada y secuenciación de los blancos de extracción (7 %) corresponde a contaminación ambiental. Es presumible que estos contaminantes ambientales también afecten a muestras degradadas. Sin embargo, no es posible confundir este tipo de contaminación con los resultados endógenos a la muestra, ya que únicamente los resultados endógenos son reproducibles.
8. La ausencia de contaminación en los blancos de PCR no es suficiente para garantizar su ausencia en las muestras. Ésta puede ser detectada mediante la comparación de los perfiles genéticos obtenidos con los de los manipuladores y de las muestras analizadas en paralelo.

9. Las principales fuentes de contaminación detectadas secuenciando los amplificados son el investigador y el personal que ha manipulado la muestra con anterioridad a su análisis en el laboratorio (antropólogos y arqueólogos). Por lo tanto la obtención de perfiles genéticos de todas aquellas personas que hayan entrado en contacto con la muestra es de suma importancia.
10. La replicación del 62,96 % de los resultados en un segundo laboratorio (LAPP-PACEA, *Université de Bordeaux 1 – CNRS, Talence, Francia*) refuerza la validez de los resultados obtenidos.

CONCLUSIONES - ESTUDIO DIACRÓNICO-POBLACIONAL

Haplotipos y Haplogrupos

11. Los haplotipos compartidos entre diferentes individuos de un mismo yacimiento (Puig de la Nau y Can Sadurní) sugieren la presencia de relaciones directas a lo largo de la línea materna o cierto grado de endogamia.
12. El haplotipo 16147T 16223T 16362C, compartido entre un individuo neolítico cardial de Can Sadurní y uno post-cardial de Sant Pau del Camp, es compatible con una hipótesis de continuidad genética entre estos dos períodos, puesto que no se ha detectado en las demás poblaciones antiguas y modernas comparadas.
13. El único haplogrupo detectado en todos los conjuntos poblacionales estudiados es el haplogrupo K, que podría representar un marcador de continuidad genética desde el Neolítico hasta la población íbera en el noreste de la Península Ibérica. Este haplogrupo se encuentra presente en muestras neolíticas de Próximo Oriente, Europa Central y Francia, apuntando a su posible asociación con la expansión neolítica.

14. En las muestras neolíticas estudiadas no se han detectado aquellos haplogrupos (J, T1 y U3) asociados a la expansión neolítica desde la genética de poblaciones actuales. Estos resultados sugieren que estos haplogrupos no estuvieron necesariamente asociados con la llegada de nueva población neolítica, poniendo en discusión las conclusiones derivadas de estudios basados en la genética de poblaciones actuales.
15. Los resultados obtenidos sugieren afinidad entre las primeras comunidades neolíticas del Levante español y la población actual de Próximo Oriente, por la presencia de haplogrupos compartidos (N* y X1), poco comunes en la población europea actual. Estos datos sugieren un movimiento poblacional desde Próximo Oriente asociado a la neolitización del noreste de la Península Ibérica.
16. En el conjunto de datos analizados, también destaca la presencia del haplogrupo U5 asociado a la cultura mesolítica centro y norte europea. Este resultado podría sugerir cierta continuidad genética con el sustrato paleolítico.

Modelos demográficos y simulaciones

17. Los modelos panmícticos no son adecuados para el estudio de los procesos demográficos ocurridos en Europa y en la Península Ibérica desde la llegada del hombre moderno hasta la actualidad, sin embargo los modelos estructurados permiten explicar los datos genéticos de manera más exhaustiva.
18. Los resultados de simulación llevados a cabo con la población neolítica antigua sugieren la llegada a principios del Neolítico de pequeños grupos genética y culturalmente en conexión con Próximo Oriente. También sugieren que durante este período la deriva genética jugó un papel importante y que el modelo que mejor se ajusta a los resultados es una colonización pionera de los primeros grupos neolíticos en el noreste de la Península Ibérica.

19. Los resultados apuntan a la presencia de cierta continuidad genética desde el Neolítico Medio hasta la actualidad, exceptuando la población íbera. Esta última presenta cierta diferenciación respecto a todas las demás poblaciones. Dado que los individuos analizados no fueron incinerados como requería la tradición funeraria de ese período, cabe pensar que pertenecieran a un grupo étnico diferenciado con una composición genética distinta. Para confirmar esta hipótesis sería necesario ampliar el conjunto muestral.

CONCLUSIONES - RELACIONES FAMILIARES

20. La repetición de las fases de extracción y amplificación de ADN para un mínimo de dos muestras por individuo y la correcta interpretación de los perfiles genéticos de nuDNA ha permitido obtener cuatro perfiles completos y tres parciales a partir de siete muestras medievales.
21. Ha sido posible determinar el sexo molecularmente en el 100 % de los casos, incluidos los individuos infantiles, para los cuales no fue posible una determinación antropométrica. Además, se ha determinado el sexo femenino del individuo 5SC, clasificado antropométricamente como varón, confirmando la concordancia entre la inscripción de la tumba (que reza *Doña Urraca García de Tapia*) y el cuerpo inhumado en su interior.
22. El análisis genético ha permitido excluir la hipótesis arqueológica de paternidad entre los individuos infantiles enterrados junto con un individuo adulto. Sin embargo, este último se ha podido relacionar con otro espécimen enterrado en un sepulcro diferente con una elevada probabilidad (99,98 %), confirmando otra hipótesis arqueológica que suponía una relación paterno-filial entre ellos, basada en inscripciones sepulcrales poco claras.

23. Existe una probabilidad menos elevada (68 %) de que los dos individuos infantiles sean hermanos o de que uno de ellos sea hijo del individuo femenino SSC. Estas probabilidades reducidas se deben a la obtención de perfiles parciales para los individuos infantiles. En estos casos la amplificación de ADN resulta menos eficiente debido a la ausencia de piezas dentales.

CONCLUSIONS

CONCLUSIONS (ENGLISH)

CONCLUSIONS – METHODOLOGY AND VALIDATION OF THE RESULTS

Experimental Methods

1. The comparative analysis of different DNA extraction methods has allowed us to select the protocol designed by Rohland and Hofreiter (2007), based on the use of silica, as a substitute for the Phenol/Chloroform based method that was commonly used in our laboratory. The former protocol is highly efficient in obtaining validated results and in eliminating inhibitors. Moreover, it does not imply exposure to toxic volatile reagents, it is scalable and it allows the overlapping of different extraction sets, the filtering of the solutions and the use of sterile materials. When working with well-preserved samples, GuSCN can be replaced by NaCl, considerably reducing the costs of the protocol.
2. It has been shown that the *Multiplex PCR kit (QIAGEN)* is an efficient protocol for aDNA amplification. Although its high sensitivity may bring with it a greater probability of amplifying contaminants when working with extremely damaged samples, it allows the analyses of inhibited DNA extracts. In this sense, the cost-benefit ratio is positive.
3. It has been possible to obtain genetic information from nuclear markers in ancient samples using some tools designed for the field of forensics, such as the *AMPFSTR® MiniFiler™ PCR Amplification Kit (Applied Biosystems)*. However, in this case amplification success depends on sample preservation. Moreover, the software has been extremely useful in accepting and/or rejecting archaeological hypotheses of kinship.

4. The methodological approach made it possible to obtain reproducible genetic information from samples that were over 7,000 years old and had come from temperate environments.

Prospection

5. Sample origin and preservation are the variables that most influence the recovery of DNA from archaeological samples. The depositional environment seems to play a more important role than sample age in DNA molecular preservation.
6. Spectrophotometric quantification of DNA by UV absorbance does not show either the amplifiable DNA content or the presence of inhibitory molecules. This is not an appropriate screening method in the field of ancient DNA.

Validation of results

7. The contamination detected through the repeated amplification of extraction blanks (7 %) corresponds to environmental DNA. These air contaminants are also expected to affect degraded samples. However, it is not possible to confuse this kind of contamination with endogenous DNA because only endogenous results are reproducible.
8. The fact that contamination in PCR blanks goes undetected is not enough to guarantee its absence in the samples. This kind of contamination can be detected by comparing genetic profiles of the samples with the ones obtained from the researchers in charge of handling them and with other samples analysed in parallel.
9. Major sources of contamination detected by PCR amplicon sequencing include the laboratory researchers and previous handlers (anthropologists and archaeologists). Therefore, obtaining handlers' profiles is of utter importance.

10. The replication of results (62.96 %) in a second laboratory (LAPP-PACE, Université de Bordeaux 1 - CNRS, Talence, France) strengthens the validity of the results obtained.

CONCLUSIONS – DIACHRONIC-POPULATIONS STUDY

Haplotypes and Haplogroups

11. Shared haplotypes among skeletons from the same site (Puig de la Nau and Can Sadurní) suggest the existence of maternal lineage relationships or a certain degree of intra-site endogamy.
12. The haplotype 16362C 16223T 16147T is shared between one Cardial Neolithic individual from Can Sadurní and one post-Cardial specimen from Sant Pau del Camp. This might reflect some genetic continuity between these two periods, since this haplotype has not been found in the ancient and modern populations compared.
13. The only mitochondrial haplogroup found in all ancient datasets is haplogroup K. This might represent a marker of genetic continuity from the Neolithic to the Iberian age in North Eastern Iberia. In addition, this haplogroup has been detected in Near Eastern Neolithic samples and in Neolithic specimens from Central Europe and France, pointing to a possible association of this marker with Neolithic expansion.
14. Haplogroups J, T1, and U3, linked to Neolithic expansion according to modern population genetics studies, have not been found in the Neolithic samples studied here. This result suggests that these haplogroups were not necessarily linked to the arrival of the Neolithic populations, which casts doubts upon the conclusions drawn from studies based on modern population genetics.
15. The results obtained suggest some affinities between North Eastern Iberian Early Neolithic communities and current Near Eastern populations, since

they share haplogroups N* and X1, both uncommon in modern European populations. These data suggest a Neolithic migration from the Near East to North Eastern Iberia.

16. Haplogroup U5, which has been associated to Central and Northern European Mesolithic culture, is also present in our ancient dataset. This result might suggest a certain degree of genetic continuity with the previous Palaeolithic substrate.

Demographic models and simulations

17. Panmictic models are not suitable for the study of the demographic processes that took place in Europe and in the Iberian Peninsula since the arrival of modern humans. Structured models explain ancient and modern genetic data more accurately.
18. Early Neolithic population simulation results suggest that small groups with genetic and cultural connections to the Near East arrived at the beginning of the Neolithic. Moreover, these results highlight the important role that genetic drift played during this period. In this framework, the model which best fit with the results obtained in North Eastern Iberia is a pioneer colonization of the first Neolithic groups.
19. Our results suggest the presence of some genetic continuity from the Middle Neolithic to present times, except for the Iberian period, which is genetically different from all other populations compared. The Iberian specimens analyzed were not incinerated, as required by the funerary tradition of that period. As a consequence, it is possible that they belonged to a distinct ethnic group with a different genetic pool. To confirm this hypothesis it would be necessary to extend our dataset.

CONCLUSIONS – KINSHIP RELATIONSHIPS

20. Four complete and three partial nuDNA profiles were obtained from seven medieval specimens. This was possible by repeating DNA extractions and amplifications for a minimum of two samples per individual and by a suitable interpretation of the genetic profiles.
21. Sex was molecularly determined in all the individuals studied, including infants, for whom an anthropometric determination was not possible. Furthermore, individual 5SC, anthropometrically classified as a male, was genetically typed as a female. This confirms the correlation between the inscription found at the tomb (which reads Mrs. Urraca García de Tapia) and the body buried inside.
22. The archaeological hypothesis of paternity between the two infants buried with an adult was ruled out through genetic analysis. However, the latter was genetically related to another specimen buried in a different tomb, with a high probability of paternity (99.98 %). This result confirms another archaeological hypothesis that assumed a father-son relationship between them based on unclear sepulchral inscriptions.
23. The sibling relationship between the two infants and the maternity of one of them by 5SC are supported by lower probabilities (68 %). Partial profiles recovered for both infants are responsible of these low probability values. The absence of dental samples in newborn individuals made DNA amplification less efficient.

BIBLIOGRAFÍA

BIBLIOGRAFÍA

- Allouche M, Hamdoum M, Mangin P, y Castella V. 2008. Genetic identification of decomposed cadavers using nails as DNA source. *Forensic Science International: Genetics* 3:46-49.
- Alonso A, Martín P, Albarrán C, García P, García O, de Simón LF, García-Hirschfeld J, Sancho M, de La Rúa C, y Fernández-Piqueras J. 2004. Real-time PCR designs to estimate nuclear and mitochondrial DNA copy number in forensic and ancient DNA studies. *Forensic Sci. Int* 139:141-149.
- Alroy J. 2001. A Multispecies Overkill Simulation of the End-Pleistocene Megafaunal Mass Extinction. *Science* 292:1893 -1896.
- Alvarez AJ, Khanna M, Toranzos GA, y Stotzky G. 1998. Amplification of DNA bound on clay minerals. *Molecular Ecology* 7:775-778.
- Alvarez García A, Muñoz I, Pestoni C, Lareu MV, Rodríguez-Calvo MS, y Carracedo A. 1996. Effect of environmental factors on PCR-DNA analysis from dental pulp. *Int. J. Legal Med* 109:125-129.
- Alvarez-Iglesias V, Mosquera-Miguel A, Cerezo M, Quintáns B, Zarrabeitia MT, Cuscó I, Lareu MV, García O, Pérez-Jurado L, Carracedo A, et al. 2009. New population and phylogenetic features of the internal variation within mitochondrial DNA macrohaplogroup R0. *PLoS ONE* 4:e5112.
- Ammerman AJ, y Cavalli-Sforza LL. 1971. Measuring the Rate of Spread of Early Farming in Europe. *Man* 6:674-688.
- Ammerman AJ, y Cavalli-Sforza LL. 1984. *The Neolithic Transition and the Genetics of Populations in Europe*. Princeton Univ Pr.
- Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, et al. 1981. Sequence and organization of the human mitochondrial genome. *Nature* 290:457-465.
- Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, y Howell N. 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 23:147.
- Arroyo-Pardo E, Ocaña MA, Arroyo JJ, Pérez-Pérez A, Turbón D, Tranco G, Rodríguez-Albarrán MS, Casas JD, Robledo B, y Bandrés F. 2002. HPLC and UV-Spectrophotometry Examination of Ancient DNA and PCR Inhibitors in Old Human Remains. *Ancient Biomolecules* 4:33.
- Balaresque P, Bowden GR, Adams SM, Leung H-Y, King TE, Rosser ZH, Goodwin J, Moisan J-P, Richard C, Millward A, et al. 2010. A Predominantly Neolithic Origin for European Paternal Lineages. *PLoS Biol* 8:e1000285.
- Balding DJ, y Buckleton J. 2009. Interpreting low template DNA profiles. *Forensic Science International: Genetics* 4:1-10.
- Bandelt H-J, Macaulay V, y Richards M eds. 2006. Estimation of Mutation Rates and Coalescence Times: Some Caveats. En: *Human Mitochondrial DNA and the Evolution of*

- Homo sapiens. Vol. 18. Springer Berlin Heidelberg.
- Bandelt H-J. 2007. Clock debate: when times are a-changin': Time dependency of molecular rate estimates: tempest in a teacup. *Heredity* 100:1-2.
- Barandiarán I, Martí Oliver B, y Rincón Martínez M. 2007. Prehistoria de la Península Ibérica. 6o ed. Editorial Ariel, S.A.
- Barbujani G, Bertorelle G, y Chikhi L. 1998. Evidence for Paleolithic and Neolithic gene flow in Europe. *Am J Hum Genet* 62:488-492.
- Barbujani G, y Bertorelle G. 2001. Genetics and the population history of Europe. *Proceedings of the National Academy of Sciences* 98:22 -25.
- Barbujani G, y Chikhi L. 2006. DNAs from the European Neolithic. *Heredity* 97:84-85.
- Barbujani G, y Goldstein DB. 2004. Africans and Asians abroad: genetic diversity in Europe. *Annu Rev Genomics Hum Genet* 5:119-150.
- Beckmann JS, Estivill X, y Antonarakis SE. 2007. Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability. *Nat. Rev. Genet* 8:639-646.
- Belle EMS, Ramakrishnan U, Mountain JL, y Barbujani G. 2006. Serial coalescent simulations suggest a weak genealogical relationship between Etruscans and modern Tuscans. *Proc. Natl. Acad. Sci. U.S.A* 103:8012-8017.
- Bermejo Tirado J. 2007. Breve historia de los íberos. Ediciones Nowtilus, S.L.
- Bermisheva M, Tambets K, Villems R, y Khushnutdinova E. 2002. Diversity of mitochondrial DNA haplotypes in ethnic populations of the Volga-Ural region of Russia. *Mol. Biol. (Mosk.)* 36:990-1001.
- Bernabeu J. 2002. The social and symbolic context of Neolithization. En: *El paisaje en el Neolítico mediterráneo*. Valencia: Saguntum-PLAV, extra 5. p 209-235.
- Bernabò Brea L 1946. *Gli Scavi nella caverna delle Arene Candide*. Bordighera: Istituto di Studi Liguri.
- Bertranpetit J, y Cavalli-Sforza LL. 1991. A genetic reconstruction of the history of the population of the Iberian Peninsula. *Ann. Hum. Genet* 55:51-67.
- Blasco A, Edo M, Villalba MJ, y Saña M. 2005. Primeros datos sobre la utilización sepulcral de la Cueva de Can Sadurní (Begues, Baix Llobregat) en el Neolítico Cardial. En: *Actas del III Congreso del Neolítico en la Península Ibérica*. Pablo Arias, Roberto Otañón, Cristina García-Moncó. Santander: Servicio de Publicaciones, Universidad de Cantabria.
- Blow MJ, Zhang T, Woyke T, Speller CF, Krivoschapkin A, Yang DY, Derevianko A, y Rubin EM. 2008. Identification of ancient remains through genomic sequencing. *Genome Res* 18:1347-1353.
- Blázquez J, Alvar J, y Wagner C. 1999. Fenicios y cartagineses en el Mediterráneo. Ediciones Cátedra, S.A.
- Bocquet-Appel J-P, y Bar-Yosef O. 2008. *The Neolithic Demographic Transition and its Consequences*. Dordrecht: Springer Netherlands.

- Bogdanowicz W, Allen M, Branicki W, Lembring M, Gajewska M, y Kupiec T. 2009. Genetic identification of putative remains of the famous astronomer Nicolaus Copernicus. *Proceedings of the National Academy of Sciences* 106:12279 -12282.
- Bollongino R, Tresset A, y Vigne J-D. 2008. Environment and excavation: Pre-lab impacts on ancient DNA analyses. *Comptes Rendus Palevol* 7:91-98.
- Bosch Gimpera P. 1919. Prehistoria catalana. En: *Enciclopedia catalana*. Vol. XVI. Editorial catalana. Barcelona.
- Bosch Gimpera P. 1932. *Etnologia de la península ibèrica*. Barcelona: Editorial Alpha.
- Bosch Gimpera P. 1940. The Types and Chronology of Western European Beakers. *Man*:6-10.
- Bouakaze C, Keyser C, Amory S, Crubézy E, y Ludes B. 2007. First successful assay of Y-SNP typing by SNaPshot minisequencing on ancient DNA. *Int. J. Legal Med* 121:493-499.
- Bouwman AS, y Brown TA. 2002. Comparison Between Silica-based Methods for the Extraction of DNA from Human Bones from 18th to Mid-19th Century London. *Ancient Biomolecules* 4:173-178.
- Bramanti B, Thomas MG, Haak W, Unterlaender M, Jores P, Tambets K, Antanaitis-Jacobs I, Haidle MN, Jankauskas R, Kind C-J, et al. 2009. Genetic Discontinuity Between Local Hunter-Gatherers and Central Europe's First Farmers. *Science* 326:137-140.
- Briggs AW, Good JM, Green RE, Krause J, Maricic T, Stenzel U, Lalueza-Fox C, Rudan P, Brajković D, Kučan Ž, et al. 2009. Targeted Retrieval and Analysis of Five Neandertal mtDNA Genomes. *Science* 325:318 -321.
- Brotherton P, Endicott P, Sanchez JJ, Beaumont M, Barnett R, Austin J, y Cooper A. 2007. Novel high-resolution characterization of ancient DNA reveals C > U-type base modification events as the sole cause of post mortem miscoding lesions. *Nucleic Acids Res* 35:5717-5728.
- Buckleton JS, Clayton T, y Triggs C. 2004. Parentage Testing. En: *Forensic DNA Evidence Interpretation*. 1o ed. CRC Press. p 349-402.
- Budowle B, Eisenberg AJ, y van Daal A. 2009. Validity of low copy number typing and applications to forensic science. *Croat. Med. J.* 50:207-217.
- Burger J, Hummel S, Hermann B, y Henke W. 1999. DNA preservation: a microsatellite-DNA study on ancient skeletal remains. *Electrophoresis* 20:1722-1728.
- Burger J, Kirchner M, Bramanti B, Haak W, y Thomas MG. 2007. Absence of the lactase-persistence-associated allele in early Neolithic Europeans. *Proceedings of the National Academy of Sciences* 104:3736 -3741.
- Busby GBJ, Brisighelli F, Sánchez-Diz P, Ramos-Luis E, Martínez-Cadenas C, Thomas MG, Bradley DG, Gusmão L, Winney B, Bodmer W, et al. 2011. The peopling of Europe and the cautionary tale of Y chromosome lineage R-M269. *Proceedings of the Royal Society B: Biological Sciences*.
- Butler JM. 2005. *Forensic DNA typing: biology, technology, and genetics of STR markers*. Academic Press.

- Cabellos T. 2008. Cova de Can Sadurní (Bergues, Baix de Llobregat). Muestras seleccionadas para el análisis de ADN. Informe Antropológico. Universidad Complutense de Madrid.
- Cai D, Tang Z, Han L, Speller CF, Yang DY, Ma X, Cao J, Zhu H, y Zhou H. 2009. Ancient DNA provides new insights into the origin of the Chinese domestic horse. *Journal of Archaeological Science* 36:835-842.
- Cann RL, Stoneking M, y Wilson AC. 1987. Mitochondrial DNA and human evolution. *Nature* 325:31-36.
- Cano RJ, Borucki MK, Higby-Schweitzer M, Poinar HN, Poinar GO Jr, y Pollard KJ. 1994. *Bacillus* DNA in fossil bees: an ancient symbiosis? *Appl. Environ. Microbiol* 60:2164-2167.
- Cano RJ, Poinar HN, Pieniazek NJ, Acra A, y Poinar GO Jr. 1993. Amplification and sequencing of DNA from a 120-135-million-year-old weevil. *Nature* 363:536-538.
- Capelli C, Onofri V, Brisighelli F, Boschi I, Scarnicci F, Masullo M, Ferri G, Tofanelli S, Tagliabracci A, Gusmao L, et al. 2009. Moors and Saracens in Europe: estimating the medieval North African male legacy in southern Europe. *Eur. J. Hum. Genet* 17:848-852.
- Caramelli D, Lalueza-Fox C, Capelli C, Lari M, Sampietro ML, Gigli E, Milani L, Pilli E, Guimaraes S, Chiarelli B, et al. 2007. Genetic analysis of the skeletal remains attributed to Francesco Petrarca. *Forensic Science International* 173:36-40.
- Caramelli D, Lalueza-Fox C, Vernesi C, Lari M, Casoli A, Mallegni F, Chiarelli B, Dupanloup I, Bertranpetit J, Barbujani G, et al. 2003. Evidence for a genetic discontinuity between Neandertals and 24,000-year-old anatomically modern Europeans. *Proc. Natl. Acad. Sci. U.S.A* 100:6593-6597.
- Cavalli-Sforza LL, Menozzi P, y Piazza A. 1994. The history and geography of human genes. Princeton University Press.
- Cavalli-Sforza LL, y Minch E. 1997. Paleolithic and Neolithic lineages in the European mitochondrial gene pool. *Am J Hum Genet* 61:247-254.
- Cherni L, Fernandes V, Pereira JB, Costa MD, Goios A, Frigi S, Yacoubi-Loueslati B, Amor MB, Slama A, Amorim A, et al. 2009. Post-last glacial maximum expansion from Iberia to North Africa revealed by fine characterization of mtDNA H haplogroup in Tunisia. *American Journal of Physical Anthropology* 139:253-260.
- Chikhi L, Destro-Bisol G, Bertorelle G, Pascali V, y Barbujani G. 1998. Clines of nuclear DNA markers suggest a largely Neolithic ancestry of the European gene pool. *Proc Natl Acad Sci U S A* 95:9053-9058.
- Chikhi L, Nichols RA, Barbujani G, y Beaumont MA. 2002. Y genetic data support the Neolithic demic diffusion model. *Proc. Natl. Acad. Sci. U.S.A* 99:11008-11013.
- Childe VG. 1930. The origin of the Bell-Beaker. *Man* XXX.
- Chilvers ER, Bouwman AS, Brown KA, Arnott RG, Prag AJNW, y Brown TA. 2008. Ancient DNA in human bones from Neolithic and Bronze Age sites in Greece and Crete. *Journal of Archaeological Science* 35:2707-2714.
- Chinnery PF. 2006. Mitochondrial DNA in Homo Sapiens. En: Bandelt H-J, Macaulay V, Richards M, editores. *Human Mitochondrial DNA and the Evolution of Homo sapiens*. Vol. 18.

Nucleic Acids and Molecular Biology. Springer Berlin Heidelberg. p 3-15.

Clark PU, Dyke AS, Shakun JD, Carlson AE, Clark J, Wohlfarth B, Mitrovica JX, Hostetler SW, y McCabe AM. 2009. The Last Glacial Maximum. *Science* 325:710 -714.

Clayton TM, Whitaker JP, Sparkes R, y Gill P. 1998. Analysis and interpretation of mixed forensic stains using DNA STR profiling. *Forensic Sci. Int* 91:55-70.

Coble MD, Loreille OM, Wadhams MJ, Edson SM, Maynard K, Meyer CE, Niederstätter H, Berger C, Berger B, Falsetti AB, et al. 2009. Mystery Solved: The Identification of the Two Missing Romanov Children Using DNA Analysis. *PLoS ONE* 4:e4838.

Cohen MN. 1984. *Paleopathology at the Origins of Agriculture*. 1o ed. Academic Press.

Collins MJ, Penkman KEH, Rohland N, Shapiro B, Dobberstein RC, Ritz-Timme S, y Hofreiter M. 2009. Is amino acid racemization a useful tool for screening for ancient DNA in bone? *Proceedings of the Royal Society B: Biological Sciences* 276:2971 -2977.

Condeemi S. 2011. *Continuity and Discontinuity in the Peopling of Europe*. Springer.

Cooper A, y Poinar HN. 2000. *Ancient DNA: Do It Right or Not at All*. *Science* 289:1139.

Cruciani F, Trombetta B, Antonelli C, Pascone R, Valesini G, Scalzi V, Vona G, Melegh B, Zagradisnik B, Assum G, et al. 2011. Strong intra- and inter-continental differentiation revealed by Y chromosome SNPs M269, U106 and U152. *Forensic Sci Int Genet* 5:e49-52.

Csányi B, Bogács-Szabó E, Tömöry G, Czibula Á, Priskin K, Csősz A, Mende B, Langó P, Csete K, Zsolnai A, et al. 2008. Y-Chromosome Analysis of Ancient Hungarian and Two Modern Hungarian-Speaking Populations from the Carpathian Basin. *Annals of Human Genetics* 72:519-534.

Curat M, Poloni E, y Sanchez-Mazas A. 2010. Human genetic differentiation across the Strait of Gibraltar. *BMC Evolutionary Biology* 10:237.

DeSalle R, Barcia M, y Wray C. 1993. PCR jumping in clones of 30-million-year-old DNA fragments from amber preserved termites (*Mastotermes electrodominicus*). *Experientia* 49:906-909.

DeSalle R, Gatesy J, Wheeler W, y Grimaldi D. 1992. DNA sequences from a fossil termite in Oligo-Miocene amber and their phylogenetic implications. *Science* 257:1933-1936.

Deguiloux M-F, Ricaud S, Leahy R, y Pemonge M-H. 2011a. Analysis of ancient human DNA and primer contamination: one step backward one step forward. *Forensic Sci. Int.* 210:102-109.

Deguiloux M-F, Soler L, Pemonge M-H, Scarre C, Jousaume R, y Laporte L. 2011b. News from the west: Ancient DNA from a French megalithic burial chamber. *Am. J. Phys. Anthropol.* 144:108-118.

Delibes de Castro G, y Fernández Miranda M. 1999. Los orígenes de la civilización: el calcolítico en el Viejo Mundo. *Síntesis*.

Dennell R. 1992. The origins of European crop agriculture. En: *The Origins of Crop Agriculture*. P. J. Watson and W. Cowan. Washington: Smithsonian Institution Publications. p 71-100.

Dolukhanov P. 1993. Foraging and farming groups in north-eastern and north-western Europe:

- identity and interaction. En: *Cultural Transformations and Interactions in Eastern Europe*. Chapman J, Dolukhanov P. Aldershot: Avebury.
- Dupanloup I, Bertorelle G, Chikhi L, y Barbujani G. 2004. Estimating the Impact of Prehistoric Admixture on the Genome of Europeans. *Molecular Biology and Evolution* 21:1361 - 1372.
- Egeland T, Mostad PF, Mevåg B, y Stenersen M. 2000. Beyond traditional paternity and identification cases. Selecting the most probable pedigree. *Forensic Sci. Int* 110:47-59.
- Eiroa JJ. 2006. *Nociones de Prehistoria general*. Editorial Ariel.
- Endicott P, y Ho SYW. 2008. A Bayesian evaluation of human mitochondrial substitution rates. *Am. J. Hum. Genet* 82:895-902.
- Ennafaa H, Cabrera VM, Abu-Amero KK, González AM, Amor MB, Bouhaha R, Dzimir N, Elgaaïed AB, y Larruga JM. 2009. Mitochondrial DNA haplogroup H structure in North Africa. *BMC Genet* 10:8.
- Estebaranz F, Fernández E, Martínez L, Gamba C, Alrousan M, Turbon D, Arroyo Pardo E, Ignasi Oms J, Pérez Pérez A, y Anfruns J. 2008. Anàlisi antropològica de les restes neolítiques de la caserna de Sant Pau (biometria, dentició, ADNA i mmicroestriació dentària). *Quarhis: Quaderns d'Arqueologia i Història de la Ciutat de Barcelona*:76-82.
- Excoffier L, Novembre J, y Schneider S. 2000. Computer note. SIMCOAL: a general coalescent program for the simulation of molecular data in interconnected populations with arbitrary demography. *Journal of Heredity* 91:506 -509.
- Excoffier L, y Lischer HEL. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* 10:564-567.
- Faerman M, Filon D, Kahila G, Greenblatt CL, Smith P, y Oppenheim A. 1995. Sex identification of archaeological human remains based on amplification of the X and Y amelogenin alleles. *Gene* 167:327-332.
- Feller EA. 2009. Evaluation of the AmpFISTRRTM MiniFiler (TM) typing Kit: Mixture and non-probative studies.
- Fernández E, Ortiz JE, Pérez-Pérez A, Prats E, Turbón D, Torres T, y Arroyo-Pardo E. 2009. Aspartic acid racemization variability in ancient human remains: implications in the prediction of ancient DNA recovery. *Journal of Archaeological Science* 36:965-972.
- Fernández E. 2005. Polimorfismos de DNA mitocondrial en poblaciones antiguas de la cuenca mediterránea.
- Findlay I, Chan F-Y, Forbes K, y Mitchelson KR. 2001. High throughput genetic diagnosis of single and small numbers of cells. *Today's Life Science* 13:40-46.
- Fisher RA. 1937. The Wave of Advance of Advantageous Genes. *Annals of Eugenics* 7:355-359.
- Fletcher D. 1963. Nuevos datos sobre las relaciones neolíticas entre las costas españolas y el Mediterráneo Oriental. En: *Homenaje a P. Bosch Gimpera*. . p 167-172.
- Fortea J, y Martí B. 1984. Consideraciones sobre los inicios del Neolítico en el Mediterráneo

español. *Zephyrus* XXXVII-XXXVIII:167-199.

Fortea J. 1973. Los complejos microlaminares y geométricos del Epipaleolítico mediterráneo español.

Fullola i Pericot J, y Nadal Lorenzo J. 2005. Introducción a la prehistoria. Editorial UOC, S.L.

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, et al. 2002. The Structure of Haplotype Blocks in the Human Genome. *Science* 296:2225 -2229.

Gamba C, Fernandez E, Oliver A, Tirado M, Baeza C, Lopezparra A, y Arroyopardo E. 2008. Population genetics and DNA preservation in ancient human remains from Eastern Spain. *Forensic Science International: Genetics Supplement Series* 1:462-464.

Gamba C, Fernández E, Tirado M, Palomo S, Deguilloux M-F, Pemonge M-H, y Arroyo-Pardo E. 2011. ADN ancien d'échantillons néolithiques de la péninsule Ibérique. En: *Bulletins de la Société Préhistorique Française. Actes du XXVIIe Congrès Préhistorique de France Transitions, Ruptures et Continuité durant la Préhistoire Transitions, Ruptures, and Continuity in Prehistory*. Bordeaux - Les Eyzies.

Gamble C, Davies W, Pettitt P, Hazelwood L, y Richards M. 2005. The Archaeological and Genetic Foundations of the European Population During the Late Glacial: Implications For. *Cambridge Archaeological Journal* 15:193-223.

Gamble C, Davies W, Pettitt P, y Richards M. 2004. Climate change and evolving human diversity in Europe during the last glacial. *Philos Trans R Soc Lond B Biol Sci* 359:243-254.

García O, Fregel R, Larruga JM, Álvarez V, Yurrebaso I, Cabrera VM, y González AM. 2011. Using mitochondrial DNA to test the hypothesis of a European post-glacial human recolonization from the Franco-Cantabrian refuge. *Heredity* 106:37-45.

García-Bour J, Pérez-Pérez A, Alvarez S, Fernández E, López-Parra AM, Arroyo-Pardo E, y Turbón D. 2004. Early population differentiation in extinct aborigines from Tierra del Fuego-Patagonia: ancient mtDNA sequences and Y-chromosome STR characterization. *Am. J. Phys. Anthropol* 123:361-370.

Garrido-Pena R. 2007. El fenómeno campaniforme: un siglo de debates sobre un enigma sin resolver. En: *Acercándonos al pasado: Prehistoria en 4 actos. El fenómeno campaniforme: un siglo de debates sobre un enigma sin resolver*. Madrid: Museo Arqueológico Nacional. Ministerio de Cultura.

Gerstenberger J, Hummel S, Schultes T, Häck B, y Herrmann B. 1999. Reconstruction of a historical genealogy by means of STR analysis and Y-haplotyping of ancient DNA. *Eur. J. Hum. Genet* 7:469-477.

Gilbert MTP, Binladen J, Miller W, Wiuf C, Willerslev E, Poinar H, Carlson JE, Leebens-Mack JH, y Schuster SC. 2007. Recharacterization of ancient DNA miscoding lesions: insights in the era of sequencing-by-synthesis. *Nucleic Acids Res* 35:1-10.

Gilbert MTP, Hansen AJ, Willerslev E, Rudbeck L, Barnes I, Lynnerup N, y Cooper A. 2003a. Characterization of Genetic Miscoding Lesions Caused by Postmortem Damage. *Am J Hum Genet* 72:48-61.

Gilbert MTP, Shapiro B, Drummond A, y Cooper A. 2005. Post-mortem DNA damage hotspots

- in Bison (*Bison bison*) provide evidence for both damage and mutational hotspots in human mitochondrial DNA. *Journal of Archaeological Science* 32:1053-1060.
- Gilbert MTP, Willerslev E, Hansen AJ, Barnes I, Rudbeck L, Lynnerup N, y Cooper A. 2003b. Distribution Patterns of Postmortem Damage in Human Mitochondrial DNA. *The American Journal of Human Genetics* 72:32-47.
- Gilbert MTP. 2006. Postmortem Damage of Mitochondrial DNA. En: Bandelt H-J, Macaulay V, Richards M, editores. *Human Mitochondrial DNA and the Evolution of Homo sapiens*. Vol. 18. Springer Berlin Heidelberg. p 91-115.
- Gill P, Ivanov PL, Kimpton C, Piercy R, Benson N, Tully G, Evett I, Hagelberg E, y Sullivan K. 1994. Identification of the remains of the Romanov family by DNA analysis. *Nat Genet* 6:130-135.
- Gill P, Puch-Solis R, y Curran J. 2009. The low-template-DNA (stochastic) thresholdâ€™Its determination relative to risk analysis for national DNA databases. *Forensic Science International: Genetics* 3:104-111.
- Gill P, Sparkes R, y Kimpton C. 1997. Development of guidelines to designate alleles using an STR multiplex system. *Forensic Sci. Int* 89:185-197.
- Gill P, Whitaker J, Flaxman C, Brown N, y Buckleton J. 2000. An investigation of the rigor of interpretation rules for STRs derived from less than 100 pg of DNA. *Forensic Sci. Int* 112:17-40.
- Gimbutas M. 1970. Proto-Indo-European Culture: The Kurgan Culture during the Fifth, Fourth, and Third Millennia B.C. En: *Indo-European and Indo-Europeans: Papers Presented at the Third Indo-European Conference at the University of Pennsylvania*. Cardona, George; Hoenigswald, Henry M.; Senn, Alfred. Philadelphia: University of Pennsylvania Press. p 155-197.
- Gjertson DW, Brenner CH, Baur MP, Carracedo A, Guidet F, Luque JA, Lessig R, Mayr WR, Pascali VL, Prinz M, et al. 2007. ISFG: Recommendations on biostatistics in paternity testing. *Forensic Science International: Genetics* 1:223-231.
- Glase JA. 1995. Validity of nucleic acid purities monitored by 260nm/280nm absorbance ratios. *BioTechniques* 18:62-63.
- Golenberg EM, Giannasi DE, Clegg MT, Smiley CJ, Durbin M, Henderson D, y Zurawski G. 1990. Chloroplast DNA sequence from a miocene *Magnolia* species. *Nature* 344:656-658.
- González AM, Brehm A, Pérez JA, Maca-Meyer N, Flores C, y Cabrera VM. 2003. Mitochondrial DNA affinities at the Atlantic fringe of Europe. *Am. J. Phys. Anthropol* 120:391-404.
- Gonçalves R, Freitas A, Branco M, Rosa A, Fernandes AT, Zhivotovsky LA, Underhill PA, Kivisild T, y Brehm A. 2005. Y-chromosome lineages from Portugal, Madeira and Açores record elements of Sephardim and Berber ancestry. *Ann. Hum. Genet* 69:443-454.
- Gray RD, y Atkinson QD. 2003. Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature* 426:435-439.
- Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MH-Y, et al. 2010. A Draft Sequence of the Neandertal Genome. *Science* 328:710 -722.

- Guilaine J. 1966. Vases Campaniformes décorés à la cordelette dans le sud de France. *Arqueología e Historia* XII:287-301.
- Guilaine J. 2003. La diffusion de l'agriculture en Europe: une hypothèse arythmique. En: *De la vague à la tombe. La conquête néolithique de la Méditerranée (8000-2000 avant J.-C.)*. Seuil. Paris. p 103-112.
- Guy Straus L. 1989. Age of the modern Europeans. *Nature* 342:476-477.
- Gérard N, Berriche S, Aouizérate A, Diéterlen F, y Lucotte G. 2006. North African Berber and Arab influences in the western Mediterranean revealed by Y-chromosome DNA haplotypes. *Hum. Biol* 78:307-316.
- Haak W, Balanovsky O, Sanchez JJ, Koshel S, Zaporozhchenko V, Adler CJ, Der Sarkissian CSI, Brandt G, Schwarz C, Nicklisch N, et al. 2010. Ancient DNA from European early neolithic farmers reveals their near eastern affinities. *PLoS Biol* 8:e1000536.
- Haak W, Brandt G, de Jong HN, Meyer C, Ganslmeier R, Heyd V, Hawkesworth C, Pike AWG, Meller H, y Alt KW. 2008. Ancient DNA, Strontium isotopes, and osteological analyses shed light on social and kinship organization of the Later Stone Age. *Proc Natl Acad Sci U S A* 105:18226-18231.
- Haak W, Forster P, Bramanti B, Matsumura S, Brandt G, Tänzer M, Villems R, Renfrew C, Gronenborn D, Alt KW, et al. 2005. Ancient DNA from the First European Farmers in 7500-Year-Old Neolithic Sites. *Science* 310:1016 -1018.
- Hagelberg E, Sykes B, y Hedges R. 1989. Ancient bone DNA amplified. *Nature* 342:485.
- Hagelberg E, y Clegg JB. 1991. Isolation and characterization of DNA from archaeological bone. *Proc. Biol. Sci* 244:45-50.
- Handt O, Höss M, Krings M, y Pääbo S. 1994. Ancient DNA: Methodological challenges. *Experientia* 50:524-529.
- Hashiyada M, Nakayashiki N, y Funayama M. 2009. Utility validation of extraction of genomic DNA from hard tissues, bone and nail, using PrepFiler™ Forensic DNA Extraction Kit. *Forensic Science International: Genetics Supplement Series* 2:172-173.
- Hawass Z, Gad YZ, Ismail S, Khairat R, Fathalla D, Hasan N, Ahmed A, Elleithy H, Ball M, Gaballah F, et al. 2010. Ancestry and Pathology in King Tutankhamun's Family. *JAMA: The Journal of the American Medical Association* 303:638 -647.
- Hedges REM, y Millard AR. 1995. Bones and Groundwater: Towards the Modelling of Diagenetic Processes. *Journal of Archaeological Science* 22:155-164.
- Herrmann B. 1994. Ancient DNA: recovery and analysis of genetic material from paleontological, archaeological, museum, medical, and forensic specimens. Springer.
- Heyer E, Zietkiewicz E, Rochowski A, Yotova V, Puymirat J, y Labuda D. 2001. Phylogenetic and familial estimates of mitochondrial substitution rates: study of control region mutations in deep-rooting pedigrees. *Am. J. Hum. Genet* 69:1113-1126.
- Higuchi R, Bowman B, Freiburger M, Ryder OA, y Wilson AC. 1984. DNA sequences from the quagga, an extinct member of the horse family. *Nature* 312:282-284.

- Ho SYW, Shapiro B, Phillips MJ, Cooper A, y Drummond AJ. 2007. Evidence for Time Dependency of Molecular Rate Estimates. *Systematic Biology* 56:515 -522.
- Hodder I. 2007. Excavating Catalhoyuk: South, North and KOPAL area reports from the 1995-99 seasons. McDonald Institute for Archaeological Research.
- Hofreiter M, Jaenicke V, Serre D, Haeseler A von, y Pääbo S. 2001a. DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Res* 29:4793-4799.
- Hofreiter M, Serre D, Poinar HN, Kuch M, y Paabo S. 2001b. Ancient DNA. *Nat Rev Genet* 2:353-359.
- Howell N, Howell C, y Elson JL. 2008. Molecular clock debate: Time dependency of molecular rate estimates for mtDNA: this is not the time for wishful thinking. *Heredity* 101:107-108.
- Hummel S. 2003. Ancient DNA typing: methods, strategies, and applications. Springer.
- Hänni C, Brousseau T, Laudet V, y Stehelin D. 1995. Isopropanol precipitation removes PCR inhibitors from ancient bone extracts. *Nucleic Acids Research* 23:881 -882.
- Höss M, Jaruga P, Zastawny TH, Dizdaroglu M, y Paabo S. 1996. DNA Damage and DNA Sequence Retrieval from Ancient Tissues. *Nucleic Acids Research* 24:1304 -1307.
- Ingman M, Kaessmann H, Paabo S, y Gyllensten U. 2000. Mitochondrial genome variation and the origin of modern humans. *Nature* 408:708-713.
- Izagirre N, y de la Rúa C. 1999. An mtDNA analysis in ancient Basque populations: implications for haplogroup V as a marker for a major paleolithic expansion from southwestern europe. *Am J Hum Genet* 65:199-207.
- Jehaes E, Pfeiffer H, Toprak K, Decorte R, Brinkmann B, y Cassiman JJ. 2001. Mitochondrial DNA analysis of the putative heart of Louis XVII, son of Louis XVI and Marie-Antoinette. *Eur. J. Hum. Genet* 9:185-190.
- Jobling MA, Hurles M, y Tyler-Smith C. 2003. *Human Evolutionary Genetics: Origins, Peoples and Disease*. 1o ed. Garland Science.
- Jobling MA, y Tyler-Smith C. 2003. The human Y chromosome: an evolutionary marker comes of age. *Nat Rev Genet* 4:598-612.
- Jordá F. 1998. *Prehistoria*. Editorial Gredos.
- Juan-Cabanilles J, y Martí B. 2002. Poblamiento y procesos culturales en la península ibérica del VII al Vmilenio a.C. En: *El paisaje en el Neolítico mediterráneo*. Valencia: Saguntum-PLAV, extra 5. p 45-89.
- Kemp BM, Malhi RS, McDonough J, Bolnick DA, Eshleman JA, Rickards O, Martinez-Labarga C, Johnson JR, Lorenz JG, Dixon EJ, et al. 2007. Genetic analysis of early holocene skeletal remains from Alaska and its implications for the settlement of the Americas. *Am. J. Phys. Anthropol* 132:605-621.
- Kemp BM, y Smith DG. 2005. Use of bleach to eliminate contaminating DNA from the surface of bones and teeth. *Forensic Sci. Int* 154:53-61.

- Keyser C, Bouakaze C, Crubézy E, Nikolaev VG, Montagnon D, Reis T, y Ludes B. 2009. Ancient DNA provides new insights into the history of south Siberian Kurgan people. *Hum Genet* 126:395-410.
- Keyser-Tracqui C, Crubézy E, y Ludes B. 2003. Nuclear and Mitochondrial DNA Analysis of a 2,000-Year-Old Necropolis in the Egyin Gol Valley of Mongolia. *Am J Hum Genet* 73:247-260.
- Kloosterman AD, y Kersbergen P. 2003. Efficacy and limits of genotyping low copy number (LCN) DNA samples by multiplex PCR of STR loci. *J. Soc. Biol* 197:351-359.
- Knapp M, y Hofreiter M. 2010. Next Generation Sequencing of Ancient DNA: Requirements, Strategies and Perspectives. *Genes* 1:227-243.
- Kolman CJ, y Tuross N. 2000. Ancient DNA analysis of human populations. *Am. J. Phys. Anthropol* 111:5-23.
- Krause J, Fu Q, Good JM, Viola B, Shunkov MV, Derevianko AP, y Paabo S. 2010. The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. *Nature* 464:894-897.
- Kricker MC, y Drake JW. 1990. Heat mutagenesis in bacteriophage T4: another walk down the transversion pathway. *J. Bacteriol* 172:3037-3039.
- Krings M, Capelli C, Tschentscher F, Geisert H, Meyer S, von Haeseler A, Grossschmidt K, Possnert G, Paunovic M, y Paabo S. 2000. A view of Neandertal genetic diversity. *Nat Genet* 26:144-146.
- Krings M, Stone A, Schmitz RW, Krainitzki H, Stoneking M, y Pääbo S. 1997. Neandertal DNA sequences and the origin of modern humans. *Cell* 90:19-30.
- Kurosaki K, Matsushita T, y Ueda S. 1993. Individual DNA identification from ancient human remains. *Am. J. Hum. Genet.* 53:638-643.
- Lacan M, Keyser C, Ricaut F-X, Brucato N, Duranthon F, Guilaine J, Crubézy E, y Ludes B. 2011. Ancient DNA reveals male diffusion through the Neolithic Mediterranean route. *Proceedings of the National Academy of Sciences* 108:9788 -9791.
- Lanting J, y van del Waals J. 1974. Beaker Culture Relations in the Lower Rhine Basin. En: *Glockenbechersimposion Oberried*. Lanting et al. Bussum-Haarlem. p 1-80.
- Lassen C, Hummel S, y Herrmann B. 1994. Comparison of DNA extraction and amplification from ancient human bone and mummified soft tissue. *Int. J. Legal Med.* 107:152-155.
- Lee HC, Pagliaro EM, Berka KM, Folk NL, Anderson DT, Ruano G, Keith TP, Phipps P, Herrin GL Jr, y Garner DD. 1991. Genetic markers in human bone: I. Deoxyribonucleic acid (DNA) analysis. *J. Forensic Sci* 36:320-330.
- Lindahl T. 1993. Instability and decay of the primary structure of DNA. *Nature* 362:709-715.
- Lindahl T. 1997. Facts and artifacts of ancient DNA. *Cell* 90:1-3.
- Loogväli E-L, Roostalu U, Malyarchuk BA, Derenko MV, Kivisild T, Metspalu E, Tambets K, Reidla M, Tolk H-V, Parik J, et al. 2004. Disuniting Uniformity: A Pied Cladistic Canvas of mtDNA Haplogroup H in Eurasia. *Molecular Biology and Evolution* 21:2012 -2021.

- Lorenz MG, y Wackernagel W. 1987. Adsorption of DNA to sand and variable degradation rates of adsorbed DNA. *Appl Environ Microbiol* 53:2948-2952.
- Lull Santiago V, González Marcén P, y Risch R. 1991. *Arqueología de Europa, 2250-1200 a.C. : una introducción a la edad del bronce*. Editorial Síntesis, S.A.
- Lynnerup N. 2007. Mummies. *American Journal of Physical Anthropology* 134:162-190.
- Malaspina P, Tsopanomichalou M, Duman T, Stefan M, Silvestri A, Rinaldi B, Garcia O, Giparaki M, Plata E, Kozlov AI, et al. 2001. A multistep process for the dispersal of a Y chromosomal lineage in the Mediterranean area. *Ann. Hum. Genet* 65:339-349.
- Mallory J. 1997. Yamna Culture. En: *Encyclopedia of Indo-European culture*. Fitzroy Dearborn.
- Malmström H, Gilbert MTP, Thomas MG, Brandström M, Storå J, Molnar P, Andersen PK, Bendixen C, Holmlund G, y Götherström A. 2009. Ancient DNA Reveals Lack of Continuity between Neolithic Hunter-Gatherers and Contemporary Scandinavians. *Current Biology* 19:1758-1762.
- Malmström H, Linderholm A, Liden K, Stora J, Molnar P, Holmlund G, Jakobsson M, y Götherström A. 2010. High frequency of lactose intolerance in a prehistoric hunter-gatherer population in northern Europe. *BMC Evolutionary Biology* 10:89.
- Malmström H, Storå J, Dalén L, Holmlund G, y Götherström A. 2005. Extensive Human DNA Contamination in Extracts from Ancient Dog Bones and Teeth. *Molecular Biology and Evolution* 22:2040-2047.
- Malyarchuk B, Derenko M, Grzybowski T, Perkova M, Rogalla U, Vanecsek T, y Tsybovsky I. 2010. The Peopling of Europe from the Mitochondrial Haplogroup U5 Perspective. *PLoS ONE* 5:e10285.
- Malyarchuk B, Rogozin I, Berikov V, y Derenko M. 2002. Analysis of phylogenetically reconstructed mutational spectra in human mitochondrial DNA control region. *Human Genetics* 111:46-53.
- Manen C, Marchand G, y Faustino Carvalho A. 2007. Le Néolithique ancien de la péninsule Ibérique : vers une nouvelle évaluation du mirage africain ? En: *Congrès du Centenaire : Un siècle de construction du discours scientifique en Préhistoire*. Vol. 3.
- Manfredi G, Thyagarajan D, Papadopoulou LC, Pallotti F, y Schon EA. 1997. The fate of human sperm-derived mtDNA in somatic cells. *Am. J. Hum. Genet* 61:953-960.
- Marota I, Basile C, Ubaldi M, y Rollo F. 2002. DNA decay rate in papyri and human remains from Egyptian archaeological sites. *Am. J. Phys. Anthropol* 117:310-318.
- Martí Oliver B. 2007. El Neolítico. En: *Prehistoria de la Península Ibérica*. 6o ed. Editorial Ariel, S.A. p 155-247.
- Maya JL. 2007. El Bronce Final y los inicios de la Edad del Hierro. En: *Prehistoria de la Península Ibérica*. 6o ed. Editorial Ariel, S.A.
- McDougall I, Brown FH, y Fleagle JG. 2005. Stratigraphic placement and age of modern humans from Kibish, Ethiopia. *Nature* 433:733-736.
- Menozi P, Piazza A, y Cavalli-Sforza L. 1978. Synthetic maps of human gene frequencies in

Europeans. *Science* 201:786-792.

Meyer S, Weiss G, y von Haeseler A. 1999. Pattern of Nucleotide Substitution and Rate Heterogeneity in the Hypervariable Regions I and II of Human mtDNA. *Genetics* 152:1103-1110.

Millar CD, Huynen L, Subramanian S, Mohandesan E, y Lambert DM. 2008. New developments in ancient genomics. *Trends Ecol. Evol. (Amst.)* 23:386-393.

Montiel Duarte R. 2001. Estudio diacrónico de la variabilidad del DNA mitocondrial en población catalana. Available from: <http://tdx.cat/handle/10803/3641>

Mulero JJ, Chang CW, Lagacé RE, Wang DY, Bas JL, McMahon TP, y Hennessy LK. 2008. Development and Validation of the Amp~~FL~~STR® MiniFiler™ PCR Amplification Kit: A MiniSTR Multiplex for the Analysis of Degraded and/or PCR Inhibited DNA. *Journal of Forensic Sciences* 53:838-852.

Muller J, y van Willigen S. 2001. New radiocarbon evidence for European Bell Beakers and the consequences for the diffusion of the Bell Beaker Phenomenon. En: *Bell beakers today: pottery, people, culture, symbols in prehistoric Europe*. Franco Nicolis. p 59-75.

Mullis KB, y Faloona FA. 1987. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Meth. Enzymol* 155:335-350.

Myres NM, Rootsi S, Lin AA, Jarve M, King RJ, Kutuev I, Cabrera VM, Khusnutdinova EK, Pshenichnov A, Yunusbayev B, et al. 2011. A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe. *Eur J Hum Genet* 19:95-101.

Nagy D, Tömöry G, Csányi B, Bogácsi-Szabó E, Czibula Á, Priskin K, Bede O, Bartosiewicz L, Downes CS, y Raskó I. 2011. Comparison of lactase persistence polymorphism in ancient and present-day Hungarian populations. *American Journal of Physical Anthropology* 145:n/a.

Neustupny J. 1963. The Bell Beaker Culture In Bohemia And Moravia. *Academic Journal Offprint From Mexico*.

Nielsen-Marsh CM, y Hedges REM. 2000. Patterns of Diagenesis in Bone I: The Effects of Site Environments. *Journal of Archaeological Science* 27:1139-1150.

Novelletto A. 2007. Y chromosome variation in Europe: continental and local processes in the formation of the extant gene pool. *Ann. Hum. Biol* 34:139-172.

Núñez C, Sosa C, Baeta M, Geppert M, Turnbough M, Phillips N, Casalod Y, Bolea M, Roby R, Budowle B, et al. 2011. Genetic analysis of 7 medieval skeletons from Aragonese Pyrenees. *Croat Med J* 52:336-343.

Oota H, Saitou N, Matsushita T, y Ueda S. 1995. A genetic study of 2,000-year-old human remains from Japan using mitochondrial DNA sequences. *Am. J. Phys. Anthropol.* 98:133-145.

Orlando L, Darlu P, Toussaint M, Bonjean D, Otte M, y Hänni C. 2006. Revisiting Neandertal diversity with a 100,000 year old mtDNA sequence. *Curr. Biol* 16:R400-402.

Otoni C, Primativo G, Hooshiar Kashani B, Achilli A, Martínez-Labarga C, Biondi G, Torroni A, y

- Rickards O. 2010. Mitochondrial haplogroup H1 in north Africa: an early holocene arrival from Iberia. *PLoS ONE* 5:e13378.
- Ovchinnikov IV, Gotherstrom A, Romanova GP, Kharitonov VM, Liden K, y Goodwin W. 2000. Molecular analysis of Neanderthal DNA from the northern Caucasus. *Nature* 404:490-493.
- van Oven M, y Kayser M. 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mutat.* 30:E386-E394.
- O'Rourke DH, Hayes MG, y Carlyle SW. 2000. Ancient DNA studies in physical anthropology. *Annual Review of Anthropology*:217-242.
- Pakendorf B, y Stoneking M. 2005. Mitochondrial DNA and human evolution. *Annu. Rev. Genom. Human Genet.* 6:165-183.
- Pala M, Achilli A, Olivieri A, Kashani BH, Perego UA, Sanna D, Metspalu E, Tambets K, Tamm E, Accetturo M, et al. 2009. Mitochondrial Haplogroup U5b3: A Distant Echo of the Epipaleolithic in Italy and the Legacy of the Early Sardinians. *The American Journal of Human Genetics* 84:814-821.
- Palanichamy M, Zhang C-L, Mitra B, Malyarchuk B, Derenko M, Chaudhuri T, y Zhang Y-P. 2010. Mitochondrial haplogroup N1a phylogeography, with implication to the origin of European farmers. *BMC Evolutionary Biology* 10:304.
- Palomino Lázaro AL, Moratinos García M, y Pastor Vásquez J. 2009. Retazos de ayer. En: *Patrimonio: Fundación del patrimonio de Castilla y León*. Vol. 37. . p 51-58.
- Pantoja JG. 2003. *Historia antigua: Grecia y Roma*. Editorial Ariel.
- Passarino G, Semino O, Magri C, Al-Zahery N, Benuzzi G, Quintana-Murci L, Andellnovic S, Bulic-Jakus F, Liu A, Arslan A, et al. 2001. The 49a,f haplotype 11 is a new marker of the EU19 lineage that traces migrations from northern regions of the black sea. *Human Immunology* 62:922-932.
- Pellicer M. 1964. *El Neolítico y el Bronce de la Cueva de la Cariguela de Piñar (Granada)*. *Trabajos de Prehistoria del Seminario de Historia Primitiva del Hombre de la Universidad de Madrid* XV.
- Pereira F, Queirós S, Gusmão L, Nijman IJ, Cuppen E, Lenstra JA, Consortium E, Davis SJM, Nejmeddine F, y Amorim A. 2009. Tracing the History of Goat Pastoralism: New Clues from Mitochondrial and Y Chromosome DNA in North Africa. *Molecular Biology and Evolution* 26:2765 -2773.
- Pereira L, Cunha C, Alves C, y Amorim A. 2005. African female heritage in Iberia: a reassessment of mtDNA lineage distribution in present times. *Hum. Biol* 77:213-229.
- Pereira V, Gomes V, Amorim A, Gusmão L, y João Prata M. 2010. Genetic characterization of uniparental lineages in populations from Southwest Iberia with past malaria endemicity. *Am. J. Hum. Biol* 22:588-595.
- Pereira V, Tomas C, Amorim A, Morling N, Gusmão L, y Prata MJ. 2011. Study of 25 X-chromosome SNPs in the Portuguese. *Forensic Sci Int Genet* 5:336-338.
- Peričić M, Lauc LB, Klarić IM, Rootsi S, Janićijević B, Rudan I, Terzić R, Čolak I, Kvesić A, Popović

- D, et al. 2005. High-Resolution Phylogenetic Analysis of Southeastern Europe Traces Major Episodes of Paternal Gene Flow Among Slavic Populations. *Molecular Biology and Evolution* 22:1964 -1975.
- Perry WL 3rd, Bass WM, Riggsby WS, y Sirotkin K. 1988. The autodegradation of deoxyribonucleic acid (DNA) in human rib bone and its relationship to the time interval since death. *J. Forensic Sci* 33:144-153.
- Plaza S, Calafell F, Helal A, Bouzerna N, Lefranc G, Bertranpetit J, y Comas D. 2003. Joining the pillars of Hercules: mtDNA sequences show multidirectional gene flow in the western Mediterranean. *Ann. Hum. Genet* 67:312-328.
- Poinar GO Jr. 1994. The range of life in amber: significance and implications in DNA studies. *Experientia* 50:536-542.
- Poinar HN, Cano RJ, y Poinar GO. 1993. DNA from an extinct plant. *Nature* 363:677.
- Poinar HN, Hofreiter M, Spaulding WG, Martin PS, Stankiewicz BA, Bland H, Evershed RP, Possnert G, y Pääbo S. 1998. Molecular Coproscopy: Dung and Diet of the Extinct Ground Sloth *Nothrotheriops shastensis*. *Science* 281:402 -406.
- Poinar HN, Höss M, Bada JL, y Pääbo S. 1996. Amino Acid Racemization and the Preservation of Ancient DNA. *Science* 272:864 -866.
- Poinar HN, Schwarz C, Qi J, Shapiro B, MacPhee RDE, Buigues B, Tikhonov A, Huson DH, Tomsho LP, Auch A, et al. 2006. Metagenomics to Paleogenomics: Large-Scale Sequencing of Mammoth DNA. *Science* 311:392 -394.
- Poinar HN. 2002. The genetic secrets some fossils hold. *Acc. Chem. Res* 35:676-684.
- Poinar HN. 2003. The top 10 list: criteria of authenticity for DNA from ancient and forensic samples. *International Congress Series* 1239:575-579.
- Price TD. 2000. *Europe's First Farmers*. Cambridge University Press.
- Prüfer K, Stenzel U, Hofreiter M, Pääbo S, Kelso J, y Green RE. 2010. Computational challenges in the analysis of ancient DNA. *Genome Biol* 11:R47.
- Pulquério MJF, y Nichols RA. 2007. Dates from the molecular clock: how wrong can we be? *Trends in Ecology & Evolution* 22:180-184.
- Pääbo S, Higuchi RG, y Wilson AC. 1989. Ancient DNA and the polymerase chain reaction. The emerging field of molecular archaeology. *J. Biol. Chem* 264:9709-9712.
- Pääbo S, Poinar H, Serre D, Jaenicke-Despres V, Hebler J, Rohland N, Kuch M, Krause J, Vigilant L, y Hofreiter M. 2004. Genetic analyses from ancient DNA. *Annu. Rev. Genet* 38:645-679.
- Pääbo S, y Wilson AC. 1988. Polymerase chain reaction reveals cloning artefacts. *Nature* 334:387-388.
- Pääbo S, y Wilson AC. 1991. Miocene DNA sequences - a dream come true? *Curr. Biol* 1:45-46.
- Pääbo S. 1985. Preservation of DNA in ancient Egyptian mummies. *Journal of Archaeological Science* 12:411-417.

- Pääbo S. 1989. Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. *Proceedings of the National Academy of Sciences* 86:1939-1943.
- Pääbo S. 1990. Amplifying ancient DNA. En: *PCR Protocols: A Guide to Methods and Applications*. London and San Diego: Academic Press. p 159-166.
- Quatrefages A, y Hamy E. 1874. La race de Cro-Magnon dans l'espace et dans le temps. *Bulletins de la Société d'Anthropologie de Paris* 9:260-266.
- Quintana-Murci L, y Fellous M. 2001. The human Y chromosome: the biological role of a «functional wasteland». *J Biomed Biotechnol* 1:18-24.
- Quintáns B, Álvarez-Iglesias V, Salas A, Lareu M, y Carracedo Á. 2004. Typing mtDNA SNPs of forensic and population interest with snapshot. *International Congress Series* 1261:419-421.
- Rasmussen M, Li Y, Lindgreen S, Pedersen JS, Albrechtsen A, Moltke I, Metspalu M, Metspalu E, Kivisild T, Gupta R, et al. 2010. Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* 463:757-762.
- Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, et al. 2006. Global variation in copy number in the human genome. *Nature* 444:444-454.
- Reidla M, Kivisild T, Metspalu E, Kaldma K, Tambets K, Tolk H, Parik J, Loogvali E, Derenko M, y Malyarchuk B. 2003. Origin and Diffusion of mtDNA Haplogroup X. *The American Journal of Human Genetics* 73:1178-1190.
- Renfrew C. 1987. *Archaeology and Language: The Puzzle of Indo-European Origins*. London: Jonathan Cape Ltd.
- Renfrew C. 1990. *Archaeology and Language: The Puzzle of Indo-European Origins*. Cambridge University Press.
- Rhouda T, Martínez-Redondo D, Gómez-Durán A, Elmtili N, Idaomar M, Díez-Sánchez C, Montoya J, López-Pérez MJ, y Ruiz-Pesini E. 2009. Moroccan mitochondrial genetic background suggests prehistoric human migrations across the Gibraltar Strait. *Mitochondrion* 9:402-407.
- Richards M, Côrte-Real H, Forster P, Macaulay V, Wilkinson-Herbots H, Demaine A, Papiha S, Hedges R, Bandelt HJ, y Sykes B. 1996. Paleolithic and neolithic lineages in the European mitochondrial gene pool. *Am. J. Hum. Genet* 59:185-203.
- Richards M, Macaulay V, Hickey E, Vega E, Sykes B, Guida V, Rengo C, Sellitto D, Cruciani F, Kivisild T, et al. 2000. Tracing European founder lineages in the Near Eastern mtDNA pool. *Am. J. Hum. Genet* 67:1251-1276.
- Richards M, Macaulay V, Sykes B, Pettitt P, Hedges R, Forster P, y Bandelt H-J. 1997. Reply to Cavalli-Sforza and Minch. *Am J Hum Genet* 61:251-254.
- Richards M, Macaulay V, Torroni A, y Bandelt H-J. 2002. In Search of Geographical Patterns in European Mitochondrial DNA. *Am J Hum Genet* 71:1168-1174.
- Richards M, y Sykes B. 1998. Reply to Barbujani et al. *The American Journal of Human Genetics* 62:491-492.

- Richards M. 2003. The Neolithic Invasion of Europe. *Annu. Rev. Anthropol.* 32:135-162.
- Richards MB, Macaulay VA, Bandelt H-J, y Sykes BC. 1998. Phylogeography of mitochondrial DNA in western Europe. *Ann Human Genet* 62:241-260.
- Rieder J, y Wartke R-B. 2009. Iron. En: *Brill's New Pauly*. Cancik, Hubert; Schneider, Helmuth. Brill.
- Rincón Martínez MÁ. 2007. El Calcolítico y la Edad del Bronce. En: *Prehistoria de la Península Ibérica*. 6o ed. Editorial Ariel, S.A.
- Robin ED, y Wong R. 1988. Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells. *J. Cell. Physiol* 136:507-513.
- Rogaev EI, Grigorenko AP, Moliaka YK, Faskhutdinova G, Goltsov A, Lahti A, Hildebrandt C, Kittler ELW, y Morozova I. 2009. Genomic identification in the historical case of the Nicholas II royal family. *Proceedings of the National Academy of Sciences* 106:5258 - 5263.
- Rohland N, y Hofreiter M. 2007a. Ancient DNA extraction from bones and teeth. *Nat. Protocols* 2:1756-1762.
- Rohland N, y Hofreiter M. 2007b. Comparison and optimization of ancient DNA extraction. *BioTechniques* 42:343-352.
- Romanowski G, Lorenz MG, y Wackernagel W. 1991. Adsorption of plasmid DNA to mineral surfaces and protection against DNase I. *Appl Environ Microbiol* 57:1057-1061.
- Roostalu U, Kutuev I, Loogväli E-L, Metspalu E, Tambets K, Reidla M, Khusnutdinova E, Usanga E, Kivisild T, y Villems R. 2007. Origin and Expansion of Haplogroup H, the Dominant Human Mitochondrial DNA Lineage in West Eurasia: The Near Eastern and Caucasian Perspective. *Molecular Biology and Evolution* 24:436 -448.
- Rootsi S, Kivisild T, Benuzzi G, Help H, Bermisheva M, Kutuev I, Barać L, Peričić M, Balanovsky O, Pshenichnov A, et al. 2004. Phylogeography of Y-Chromosome Haplogroup I Reveals Distinct Domains of Prehistoric Gene Flow in Europe. *The American Journal of Human Genetics* 75:128-137.
- Rosser ZH, Zerjal T, Hurles ME, Adojaan M, Alavantic D, Amorim A, Amos W, Armenteros M, Arroyo E, Barbujani G, et al. 2000. Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. *Am. J. Hum. Genet* 67:1526-1543.
- Rowley-Conwy P. 2009. Human Prehistory: Hunting for the Earliest Farmers. *Current Biology* 19:R948-R949.
- Sambrook J, y Russell DW. 2001. *Molecular cloning: a laboratory manual*. CSHL Press.
- Sampietro M., Lao O, Caramelli D, Lari M, Pou R, Martí M, Bertranpetit J, y Lalueza-Fox C. 2007. Palaeogenetic evidence supports a dual model of Neolithic spreading into Europe. *Proc Biol Sci* 274:2161-2167.
- Sampietro ML, Caramelli D, Lao O, Calafell F, Comas D, Lari M, Agustí B, Bertranpetit J, y Lalueza-Fox C. 2005. The Genetics of the Pre-Roman Iberian Peninsula: A mtDNA Study of Ancient Iberians. *Annals of Human Genetics* 69:535-548.

- San Valero Aparisi J. 1948. La Peninsula hispanica en el mundo neolitico. En: Seminario de Historia Primitiva del Hombre. Madrid.
- Sangmeister E. 1963. La civilisation du vase campaniforme. En: Actes du Premier Colloque Atlantique (Brest, 1961): Les civilisations atlantiques du néolithique á l'Age du Fer. Rennes. p 25-26.
- Schaffner SF. 2004. The X chromosome in population genetics. *Nat Rev Genet* 5:43-51.
- Schlumbaum A, Campos PF, Volken S, Volken M, Hafner A, y Schibler J. 2010. Ancient DNA, a Neolithic legging from the Swiss Alps and the early history of goat. *Journal of Archaeological Science* 37:1247-1251.
- Schonberg A, Theunert C, Li M, Stoneking M, y Nasidze I. 2011. High-throughput sequencing of complete human mtDNA genomes from the Caucasus and West Asia: high diversity and demographic inferences. *Eur J Hum Genet*.
- Schwartz M, y Vissing J. 2002. Paternal inheritance of mitochondrial DNA. *N. Engl. J. Med* 347:576-580.
- Schwartz M, y Vissing J. 2003. New patterns of inheritance in mitochondrial disease. *Biochem. Biophys. Res. Commun* 310:247-251.
- Schwartz TR, Schwartz EA, Mieszerski L, McNally L, y Kobilinsky L. 1991. Characterization of deoxyribonucleic acid (DNA) obtained from teeth subjected to various environmental conditions. *J. Forensic Sci* 36:979-990.
- Scozzari R, Cruciani F, Pangrazio A, Santolamazza P, Vona G, Moral P, Latini V, Varesi L, Memmi MM, Romano V, et al. 2001. Human Y-chromosome variation in the Western Mediterranean area: implications for the peopling of the region. *Human Immunology* 62:871-884.
- Semino O, Magri C, Benuzzi G, Lin AA, Al-Zahery N, Battaglia V, Maccioni L, Triantaphyllidis C, Shen P, Oefner PJ, et al. 2004. Origin, Diffusion, and Differentiation of Y-Chromosome Haplogroups E and J: Inferences on the Neolithization of Europe and Later Migratory Events in the Mediterranean Area. *Am J Hum Genet* 74:1023-1034.
- Semino O, Passarino G, Brega A, Fellous M, y Santachiara-Benerecetti AS. 1996. A view of the neolithic demic diffusion in Europe through two Y chromosome-specific markers. *Am J Hum Genet* 59:964-968.
- Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, Beckman LE, De Benedictis G, Francalacci P, Kouvatsi A, Limborska S, et al. 2000. The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: a Y chromosome perspective. *Science* 290:1155-1159.
- Serre D, Langaney A, Chech M, Teschler-Nicola M, Paunovic M, Menecier P, Hofreiter M, Possnert G, y Pääbo S. 2004. No Evidence of Neandertal mtDNA Contribution to Early Modern Humans. *PLoS Biol* 2:e57.
- Sewell J, Quinones I, Ames C, Multaney B, Curtis S, Seeboruth H, Moore S, y Daniel B. 2008. Recovery of DNA and fingerprints from touched documents. *Forensic Science International: Genetics* 2:281-285.

- Shoubridge EA. 2001. Nuclear genetic defects of oxidative phosphorylation. *Human Molecular Genetics* 10:2277-2284.
- Simoni L, Calafell F, Pettener D, Bertranpetit J, y Barbujani G. 2000a. Geographic Patterns of mtDNA Diversity in Europe. *The American Journal of Human Genetics* 66:262-278.
- Simoni L, Calafell F, Pettener D, Bertranpetit J, y Barbujani G. 2000b. Reconstruction of Prehistory on the Basis of Genetic Data. *Am J Hum Genet* 66:1177-1179.
- Smith CI, Chamberlain AT, Riley MS, Cooper A, Stringer CB, y Collins MJ. 2001. Neanderthal DNA: Not just old but old and cold? *Nature* 410:771-772.
- Smith CI, Chamberlain AT, Riley MS, Stringer C, y Collins MJ. 2003. The thermal history of human fossils and the likelihood of successful DNA amplification. *J. Hum. Evol* 45:203-217.
- Soares P, Achilli A, Semino O, Davies W, Macaulay V, Bandelt H-J, Torroni A, y Richards MB. 2010. The archaeogenetics of Europe. *Curr. Biol* 20:R174-183.
- Sokal RR, Oden NL, y Wilson C. 1991. Genetic evidence for the spread of agriculture in Europe by demic diffusion. *Nature* 351:143-145.
- Stankiewicz BA, Poinar HN, Briggs DEG, Evershed RP, y Poinar GO. 1998. Chemical preservation of plants and insects in natural resins. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 265:641-647.
- Steele J, Adams J, y Sluckin T. 1998. Modelling Paleoindian dispersals. *World Archaeol* 30:286-305.
- Steinlechner M, Berger B, Niederstätter H, y Parson W. 2002. Rare failures in the amelogenin sex test. *Int. J. Legal Med* 116:117-120.
- Stoneking M, Sherry ST, Redd AJ, y Vigilant L. 1992. New Approaches to Dating Suggest a Recent Age for the Human mtDNA Ancestor. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 337:167-175.
- Stoneking M, y Krause J. 2011. Learning about human population history from ancient and modern genomes. *Nat. Rev. Genet* 12:603-614.
- Stoneking M. 2000. Hypervariable Sites in the mtDNA Control Region Are Mutational Hotspots. *Am J Hum Genet* 67:1029-1032.
- Stringer C, y Andrews P. 2005. *The Complete World of Human Evolution*. First Edition. Thames & Hudson.
- Svoboda JA. 2007. The Gravettian on the Middle Danube. *PALEO* 19:203-220.
- Swinfield CE, Graham EAM, Nuttall D, Maguire S, Kemp A, y Ruttly GN. 2009. The use of DNA stabilizing solution to enable room temperature storage and transportation of buccal and trace sample swabs. *Forensic Science International: Genetics Supplement Series* 2:183-184.
- Tambets K, Rootsi S, Kivisild T, Help H, Serk P, Loogväli E-L, Tolk H-V, Reidla M, Metspalu E, Pliss L, et al. 2004. The Western and Eastern Roots of the Saami—the Story of Genetic «Outliers» Told by Mitochondrial DNA and Y Chromosomes. *Am J Hum Genet* 74:661-

682.

- Tarradell M. 1965. Una hipótesis que se desvanece: El papel de África en las raíces de los pueblos hispánicos. En: Homenaje a Jaime Vicens Vives. Vol. 1. Barcelona. p 173-181.
- Terberger T, y Street M. 2002. Hiatus or continuity? New results for the question of pleniglacial settlement in Central Europe. *Antiquity* 76:691-698.
- Tomas C, Sanchez J, Barbaro A, Brandt-Casadevall C, Hernandez A, Ben Dhiab M, Ramon M, y Morling N. 2008. X-chromosome SNP analyses in 11 human Mediterranean populations show a high overall genetic homogeneity except in North-west Africans (Moroccans). *BMC Evolutionary Biology* 8:75.
- de Torres T, Ortiz JE, Llamas FJ, Canoira L, Julia R, y Garcia-Martinez MJ. 2002. Bear dentine aspartic acid racemization analysis: a proxy for the dating of Pleistocene cave infills. *Archaeometry* 44:417-426.
- Torroni A, Achilli A, Macaulay V, Richards M, y Bandelt H-J. 2006. Harvesting the fruit of the human mtDNA tree. *Trends Genet* 22:339-345.
- Torroni A, Bandelt HJ, D'Urbano L, Lahermo P, Moral P, Sellitto D, Rengo C, Forster P, Savontaus ML, Bonn -Tamir B, et al. 1998. mtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe. *Am J Hum Genet* 62:1137-1152.
- Torroni A, Bandelt HJ, Macaulay V, Richards M, Cruciani F, Rengo C, Martinez-Cabrera V, Villems R, Kivisild T, Metspalu E, et al. 2001. A signal, from human mtDNA, of postglacial recolonization in Europe. *Am. J. Hum. Genet* 69:844-852.
- Torroni A, Richards M, Macaulay V, Forster P, Villems R, Norby S, Savontaus M, Huoponen K, Scozzari R, y Bandelt H. 2000. mtDNA Haplogroups and Frequency Patterns in Europe. *The American Journal of Human Genetics* 66:1173-1177.
- Torroni A, Schurr TG, Cabell MF, Brown MD, Neel JV, Larsen M, Smith DG, Vullo CM, y Wallace DC. 1993. Asian affinities and continental radiation of the four founding Native American mtDNAs. *Am. J. Hum. Genet* 53:563-590.
- Treinen F. 1970. Les poteries campaniformes en France. *Gallia Pr histoire* XIII:53-107, 263-332.
- Tully LA, Parsons TJ, Steighner RJ, Holland MM, Marino MA, y Prenger VL. 2000. A sensitive denaturing gradient-Gel electrophoresis assay reveals a high frequency of heteroplasmy in hypervariable region 1 of the human mtDNA control region. *Am. J. Hum. Genet* 67:432-443.
- Tuross N. 1994. The biochemistry of ancient DNA in bone. *Experientia* 50:530-535.
- Utrilla M del P, Lorenzo JI, Baldellou V, Sopena MC, y Ayuso P. 2008. Enterramiento masculino en fosa, cubierto de cantos rodados, en el neol tico antiguo de la cueva de Chaves. En: IV Congreso del Neol tico Peninsular. Vol. 2. Alicante: Museo Arqueol gico de Alicante - MARQ. p 131-140.
- Vanek D, Saskova L, y Koch H. 2009. Kinship and Y-Chromosome Analysis of 7th Century Human Remains: Novel DNA Extraction and Typing Procedure for Ancient Material. *Croat Med J* 50:286-295.

- Vasan S, Zhang X, Zhang X, Kapurniotu A, Bernhagen J, Teichberg S, Basgen J, Wagle D, Shih D, Terlecky I, et al. 1996. An agent cleaving glucose-derived protein crosslinks in vitro and in vivo. *Nature* 382:275-278.
- Vass AA. 2001. Beyond the grave – understanding human decomposition Arpad A. Vass. *Microbiology Today* 28:190-192.
- Vaňharová M, y Drozdová E. 2008. Sex determination of skeletal remains of 4000 year old children and juveniles from Hoštice 1 za Hanou (Czech Republic) by ancient DNA analysis. *Anthropological Review* 71:63-70.
- Vega Toscano L, Bernabeu J, y Chapa Brunet T. 2003. La prehistoria. Editorial Síntesis, S.A.
- Vernesi C, Di Benedetto G, Caramelli D, Secchieri E, Simoni L, Katti E, Malaspina P, Novelletto A, Marin VT, y Barbujani G. 2001. Genetic characterization of the body attributed to the evangelist Luke. *Proc. Natl. Acad. Sci. U.S.A* 98:13460-13463.
- Wakeley J. 1993. Substitution rate variation among sites in hypervariable region 1 of human mitochondrial DNA. *J Mol Evol* 37.
- Waldbaum JC. 1978. From Bronze to Iron: The Transition from the Bronze Age to the Iron Age in the Eastern Mediterranean. *Studies in Mediterranean Archaeology* LIV.
- Wallin IE. 1923. The Mitochondria Problem. *The American Naturalist* 57:255-261.
- Whitaker JP, Cotton EA, y Gill P. 2001. A comparison of the characteristics of profiles produced with the AMPFISTR SGM Plus multiplex system for both standard and low copy number (LCN) STR DNA analysis. *Forensic Sci. Int* 123:215-223.
- Whitehouse D, y Whitehouse R. 1977. *Archaeological Atlas of the World*. New edition. Thames & Hudson Ltd.
- Willerslev E, Hansen AJ, Binladen J, Brand TB, Gilbert MTP, Shapiro B, Bunce M, Wiuf C, Gilichinsky DA, y Cooper A. 2003. Diverse Plant and Animal Genetic Records from Holocene and Pleistocene Sediments. *Science* 300:791 -795.
- Willerslev E, y Cooper A. 2005. Ancient DNA. *Proc Biol Sci* 272:3-16.
- Wilson IG. 1997. Inhibition and facilitation of nucleic acid amplification. *Appl Environ Microbiol* 63:3741-3751.
- Wong KK, deLeeuw RJ, Dosanjh NS, Kimm LR, Cheng Z, Horsman DE, MacAulay C, Ng RT, Brown CJ, Eichler EE, et al. 2007. A Comprehensive Analysis of Common Copy-Number Variations in the Human Genome. *Am J Hum Genet* 80:91-104.
- Woodward, Weyand N, y Bunnell M. 1994. DNA sequence from Cretaceous period bone fragments. *Science* 266:1229 -1232.
- Xu H, Zhao Y, Liu Z, Zhu W, Zhou Y, y Zhao Z. 2011. Bisulfite genomic sequencing of DNA from dried blood spot microvolume samples. *Forensic Science International: Genetics* In Press, Corrected Proof.
- Yang DY, Eng B, Wayne JS, Dudar JC, y Saunders SR. 1998. Technical note: improved DNA extraction from ancient bones using silica-based spin columns. *Am. J. Phys. Anthropol* 105:539-543.

- Yeates C, Gillings MR, Davison AD, Altavilla N, y Veal DA. 1998. Methods for microbial DNA extraction from soil for PCR amplification. *Biol Proced Online* 1:40-47.
- Yousten AA, y Rippere KE. 1997. DNA similarity analysis of a putative ancient bacterial isolate obtained from amber. *FEMS Microbiology Letters* 152:345-347.
- Zalloua PA, Platt DE, El Sibai M, Khalife J, Makhoul N, Haber M, Xue Y, Izaabel H, Bosch E, Adams SM, et al. 2008. Identifying genetic traces of historical expansions: Phoenician footprints in the Mediterranean. *Am. J. Hum. Genet* 83:633-642.
- Zerjal T, Xue Y, Bertorelle G, Wells RS, Bao W, Zhu S, Qamar R, Ayub Q, Mohyuddin A, Fu S, et al. 2003. The Genetic Legacy of the Mongols. *Am J Hum Genet* 72:717-721.
- Zhivotovsky LA, Underhill PA, Cinnioglu C, Kayser M, Morar B, Kivisild T, Scozzari R, Cruciani F, Destro-Bisol G, Spedini G, et al. 2004. The Effective Mutation Rate at Y Chromosome Short Tandem Repeats, with Application to Human Population-Divergence Time. *Am J Hum Genet* 74:50-61.
- Zilhão J. 2000. From the Mesolithic to the Neolithic in the Iberian Peninsula. En: *Europe's first farmers*. Price, T. D. Cambridge: Cambridge University Press. p 144-182.
- Zilhão J. 2001. Radiocarbon evidence for maritime pioneer colonization at the origins of farming in west Mediterranean Europe. *Proceedings of the National Academy of Sciences* 98:14180 -14185.
- Zischler H, Geisert H, von Haeseler A, y Pääbo S. 1995. A nuclear «fossil» of the mitochondrial D-loop and the origin of modern humans. *Nature* 378:489-492.
- Zvelebil M. 1980. The rise of the nomads in Central Asia. En: *Cambridge Encyclopedia of Archaeology*. A. Sherratt. Cambridge: Cambridge University Press. p 320-3.
- Zvelebil M. 2000. The social context of the agricultural transition in Europe. En: *Archaeogenetics*. Colin Renfrew and Katie Boyle. Cambridge: McDonald Institute. p 57-79.
- Zvelebil M. 2001. The agricultural transition and the origins of Neolithic society in Europe. *Documenta Praehistorica XXVIII*:1-29.

CD

Contenido del CD

Alineamientos de las secuencias clonadas correspondientes a los Artículos 3 y 4.

CD content

Cloned sequences alignments correspondent to the Articles 3 and 4.